Puzzling structures enlightened

Bultema, Jelle Bernard

Puzzling structures enlightened

A structural survey of large protein complexes

Jelle Bultema

Cover: This photograph of sunlight breaking through clouds was taken by Marcel Bokhove at Hastière (Belgium) on an autumnal day.

Printed by: Ipskamp Drukkers B.V.

ISBN: 978-90-367-4372-3 printed version
ISBN: 978-90-367-4373-0 digital version

This Ph.D. study was carried out in the Biophysical Department of the Groningen Biomolecular Science and Biotechnology Insitute (GBB) of the University of Groningen, The Netherlands, and was financially supported by The Netherlands Organisation for Scientific Research (NWO).

rijksuniversiteit
groningen

Puzzling structures enlightened

A structural survey of large protein complexes

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
maandag 14 juni 2010
om 13:15 uur

door

Jelle Bernard Bultema
geboren op 13 augustus 1980
te Delfzijl

Promotor:	Prof. dr. E. J. Boekema
Beoordelingscommissie:	Prof. dr. H. P. Braun
	Prof. dr. R. Croce
	Prof. dr. B. W. Dijkstra

Table of contents

Aim and outline

Aim and outline

This thesis describes the results of structural characterization of large proteins complexes by single particle electron microscopy. The main aim of this project was to determine the structures of multi-protein complexes and to localize subunit positions within these multi-protein complexes. This information provides insights to better understand protein-protein interaction within complexes. The obtained results describe novel multi-protein complex structures.

Chapter 1 - Introduction

The first chapter gives general descriptions about the three multi-protein complexes that were examined. The chapter ends with a brief description concerning the main analysis method used for that particular research study.

Chapter 2 - Megacomplex organization of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato

The second chapter describes the structural characterization of stable mitochondrial supercomplexes which were isolated from potato tubers. A number of projection maps of such supercomplexes could be analysed. Some of them are likely breakdown products of a hypothetical complex that is organized into a string. Based on our assignment of individual complexes in the supercomplexes we propose a structural model for the largest supercomplex. The presented model of the respiratory string is a repetition of this proposed supercomplex and is in agreement with all projections of the identified supercomplexes.

Chapter 3 - The vesicle-inducing protein 1 (Vipp1) from *Synechocystis* PCC 6803 organizes into diverse higher-ordered ring structures

The third chapter describes the subcellular localization and the oligomeric organization of Vipp1 from the cyanobacterium *Synechocystis* PCC 6803. We show, by using different biochemical and biophysical experiments, that Vipp1 forms stable dimers and higher-ordered oligomers. Vipp1 is localized in the cytoplasm as well as at the cytoplasmic and thylakoid membrane. The higher-ordered oligomers are organized in ring structures with a variable diameters and molecular masses. Six different types of rings are present with an

unusual 12–17-fold symmetrical conformation. The simultaneous existence of multiple types of rings is extraordinary and suggests a special function of Vipp1.

Chapter 4 - Vipp1 and PspA: Related but not twins

After the initial characterization of Vipp1 from the cyanobacterium *Synechocystis* PCC 6803, a comparison with its bacterial homologue Phage Shock Protein A (PspA) is described. Both proteins have similar secondary structures and also form large homooligomeric rings with high molecular masses but their ring dimensions differ significantly. The structural differences such as rotational symmetry are compared. Furthermore, we suggest a spatial structural model of the observed Vipp1 rings.

Chapter 5 - Structural basis for CRISPR RNA-guided recognition of DNA by Cascade

Chapter five describes the biochemical and biophysical characterization of a ribonucleoprotein complex called Cascade. This multi-protein complex is part of the CRISPR immune system in prokaryotes which utilizes small guide RNAs to neutralize invading viruses and plasmids. We present the composition, protein stoichiometry and low-resolution structure of Cascade and show how it recognizes double-stranded DNA targets. The structure of Cascade reveals an unusual seahorse-shape that undergoes conformational changes upon target DNA binding. A structural model is presented providing insight into the molecular basis of crRNA-guided target DNA recognition.

Chapter 1

Introduction

Preface

In this thesis, three different topics are described. At first sight, these three topics, (i) respiratorial supercomplexes of mitochondria, (ii) the Vipp1 protein of cyanobacteria and (iii) the prokaryotic Cascade protein complex involved in immunity, share no resemblance at all. However, these protein complexes all have a direct or indirect relationship with oxygen. This similitude will be introduced in the coming paragraphs and used throughout the rest of the introduction.

The oxygen soap

In 1772 the Swedish chemist Carl Wilhelm Scheele discovered oxygen, which he called fire air. However, he couldn't get his work published until the year 1777, when his book *Chemische Abhandlung von der Luft und dem Feuer* came out. Meanwhile, in 1774, a British clergyman called Joseph Priestley, who wasn't familiar with the work done by Scheele, did a series of experiments and also discovered oxygen. He published his findings in 1774 and therefore he was the first one to describe oxygen, although he used the term dephlogisticated air. In 1778, the French chemist Antoine Laurent Lavoisier published *General Considerations on the Nature of Acids* were he claimed that he independently discovered oxygen. This can't be true because Priestley visited Lavoisier in October 1774 and they talked about the experiment how to get the dephlogisticated air. Furthermore, Scheele described in a letter to Lavoisier he mailed in September 1774 his discovery of a new unknown compound, but Lavoisier never confirmed he received this letter. Scheele's contribution, however, is less than trivial, because Lavoisier gave Scheele's fire air the name as we all know it today: oxygène. This is Greek for acid (*oxys*) and producer (*genēs*) because Lavoisier thought that it was part of all acids, which in the end was proved to be wrong. Actually, this description fitted better to hydrogen, but things went differently. English scientists were not so happy with the name oxygen, because to them Priestley was the first person that isolated and published about oxygen. Therefore, he should have been the one to give the name to the new element. However, it was already too late, the name was established. Especially the poem titled *Oxygen* written by Erasmus Darwin (the grandfather of Charles Darwin) and part of the popular book *The Botanic Garden* (1781) closed the door for changing the (English) name for oxygen. In the end, it was not so bad that the English people couldn't change the name of oxygen because this would not have given justice to Scheele who was the first to discover oxygen and no one else.

Oxygen and Life

Since Scheele's discovery of oxygen (O_2) scientific knowledge increased tremendously and nowadays we know many details about oxygen's crucial role for life. Most organisms require oxygen in oxidative reactions to live. Some of them use simple diffusion for oxygen uptake, whereas others have internal tubes through which air can freely diffuse, or through which it is actively pumped around (trachea). Higher organisms have special organs (lungs, gills or skin) for oxygen uptake together with a highly branched distribution system (cardiovascular system). After uptake, the oxygen will end up in a special compartment of the cell called mitochondrion. Here, together with a carbon source, it will be stepwise converted into carbon dioxide (CO_2) and water (H_2O). The structures of protein complexes involved in this conversion of oxygen are described in chapter 2. Most importantly, breakdown of oxygen is coupled to storage of chemically fixed energy, for instance in the form of adenosine-5'-triphosphate (ATP). This type of energy enables most organisms to live, which means to let take place all unfavourable or endergonic reactions.

However, without any recycling, the stock of oxygen and suitable carbon sources would deplete quickly and life, as we know it, would abruptly come to an end. Fortunately, nature found its way to overcome this problem in photosynthesis. Photosynthetic reactions take place in the membranes of microorganisms like algae and cyanobacteria, whereas higher organisms have a special cell compartment called the chloroplast for the same purpose. A structural study about a protein that is somehow involved in shaping the membranes within cyanobacteria and chloroplasts is described in chapters 3 and 4. The microorganisms and chloroplasts capture energy from light and use this energy to convert CO_2 and H_2O (produced by the mitochondria mentioned before) into carbohydrates and O_2 and thereby complete the oxygen cycle.

Oxygen and Death

Everybody knows that lack of oxygen can kill you, but also high concentrations of oxygen are damaging. Healthy persons that, for a short period, are exposed to high concentrations of oxygen will suffer from disorientation, breathing problems, and vision changes. The long-term effects are more serious: cell damage, collision of the alveoli in the lungs, retinal detachment or even death.

The last phenomenon, cell death due to exposure to high concentrations of oxygen, has been utilized by nature, such that humans (and all other animals) are able to kill invaders of their body with the use of oxygen. A white blood cell recognizes the invader as

harmful and ingests it within a membrane-bounded structure called phagosome. Activated white blood cells (or macrophages) will produce a number of antimicrobial and toxic molecules that can destroy the engulfed microorganism. Many of these toxic molecules are reactive forms of oxygen and originate from a metabolic process known as the respiratory burst. This process activates a phagosomal membrane protein complex, which reduces oxygen to superoxide ($O2^-$), an extremely reactive oxygen intermediate that is very toxic for ingested microbial invaders. Since superoxide is very reactive, it produces a whole spectrum of other toxic compounds inside the phagosome, like hydroxyl radicals and hydrogen peroxide. In specific occasions, the macrophage starts to express nitric oxide synthetase. This enzyme catalyzes a reaction between arginine and oxygen whereby nitric oxide (NO) is one of the products. Nitric oxide is also very toxic and extremely reactive and, combined with superoxide, it can result in even more antimicrobial compounds (Knight, 2000; Russell, 1995). Altogether, the potentially harmful invader will be destroyed, partially due to the oxygen-dependent killing mechanism. However, at the level of invasive DNA, oxygen is less effective and other defense tricks are in demand. Chapter 5 describes a study on a protein complex, the Cascade protein complex, that is involved in killing invading DNA in prokaryotes without the use of oxygen.

Scheele's other discoveries

Apart from the discovery of oxygen, Scheele can be held responsible for the fact that Sweden is one of the main producers of matches, because he found a way to produce large quantities of phosphorus. Next to that, he discovered several other elements, like barium, manganese, molybdenum, and tungsten.

The last two elements are presently quite useful to enhance the contrast in biological electron microscopy specimens. For practical reasons they can't be used in their pure form but are applied as salts, dissolved in water or buffer. These stain solutions are mixed with large macromolecules in a technique called negative staining. In negative staining, a thin layer of stain is allowed to dry in on a support film on which macromolecules have been absorbed. In practice, ammonium molybdate and methylamine tungstate have proven to give superior results of stain embedding. Together with uranyl acetate (a salt of uranium that was discovered by the German chemist Martin Heinrich Klaproth in 1781) they greatly enhance the contrast in standard samples of large protein complexes. The negative staining technique has been used in all studies presented in this thesis.

Mitochondria, respiratorial complexes and oxygen

The main eukaryotic cell compartment for ATP production is the mitochondrion. This organelle has two membranes, as shown in figure 1. The outer membrane is easily permeable for small molecules and ions, which is facilitated by a group of integral membrane proteins called porins. The inner membrane is impermeable for almost all small molecules and ions. Transport of these compounds across this membrane is only possible by specific transporters. The two membranes divide the space inside the mitochondrion into two aqueous compartments, the inter membrane space (IMS) and the matrix (Figure 1). The composition of small molecules and ions of the IMS is similar to that of the cytoplasm due to the permeability of the outer membrane. Larger compounds are translocated to the IMS by specific membrane-bound proteins that recognize specific targeting and sorting sequences of their substrates. The same applies for proteins within the matrix, the aqueous compartment enclosed by the inner membrane. Protein translocation is an important process, because most mitochondrial proteins are encoded by the nuclear genome and synthesized in the cytoplasm by cytoplasmic ribosomes Therefore they require translocation over both the outer and the inner membrane. The matrix is a compartment where multiple processes take place, for instance the citric acid cycle and the oxidation of fatty acids and amino acids. Many highly complex proteins are present, like the pyruvate dehydrogenase complex with its 60 subunits (Mattevi *et al*., 1992).

As mentioned before, the inner membrane is impermeable to most small molecules and ions and in particular to protons. The transport of protons across the inner membrane into the matrix is carried out by three respiratory protein complexes (NADH dehydrogenase - complex I, dimeric cytochrome *c* reductase - complex III_2, cytochrome *c* oxidase - complex IV) leading to an accumulation of protons within the matrix. The resulting electrochemical proton gradient is utilized by ATP synthase (referred to as complex V) to synthesize ATP from ADP and Pi. Proton translocation depends on the transfer of electrons from NADH or $FADH_2$, *via* complexes I and III and succinate dehydrogenase (complex II), to complex IV, where oxygen is reduced to water. Electron flow between the complexes is mediated by two mobile electron carriers, ubiquinone and cytochrome *c*, which are located in the inner membrane and IMS, respectively (Saraste, 1999).

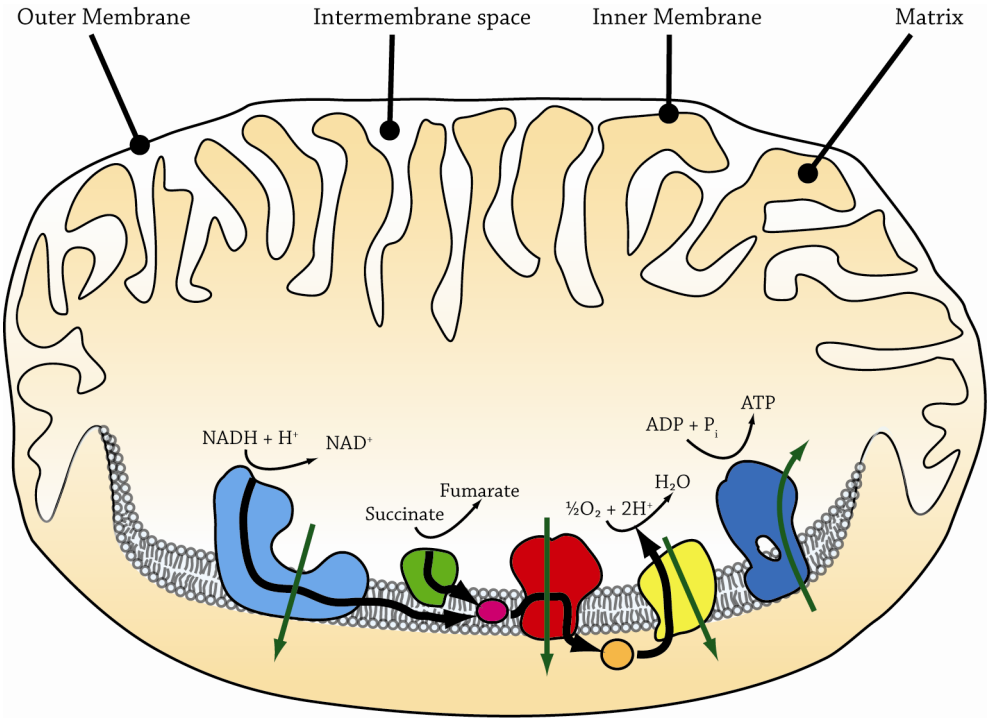


Figure 1. Structural features of the mitochondrion with a scheme of the respiratory chain. Typical mitochondria consist of an outer and inner membrane, but the size, shape and degree of convolution of the inner membrane varies among organisms. The space between the two membranes is called inter membrane space (IMS), whereas the space enveloped by the inner membrane is known as matrix. The heavily folded inner membrane provides a very large surface area that is many times larger than that of the outer membrane. The respiratory chain protein complexes are located in the inner membrane and consist of complex I (light blue), complex II (green), dimeric complex III (red), complex IV (yellow), and complex V (dark blue). Quinone (purple) and cytochrome *c* (orange) are two small mobile electron carriers that mediate electron flow (thick black arrows) between the complexes. Green arrows indicate the proton translocation direction as performed by the complexes.

Models for the overall organisation of the oxidative phosphorylation system

The five complexes together form the oxidative phosphorylation (OXPHOS) system. The overall organization of these complexes in the inner membrane has been explained by two models: the "fluid-state model" and the "solid-state model". The first model, the fluid-state model, states that the OXPHOS complexes diffuse freely in the inner membrane, and that random collision of the single complexes results into electron transfer. This model is supported by the finding that all individual OXPHOS complexes can be acquired as homogeneous and enzymatically active units (see reviews Hackenbrock *et al.*, 1986; Lenaz *et al.*, 2009). In contrast, the solid-state model postulates that specific OXPHOS complexes have stable interactions with other OXPHOS complexes, leading to the presence of so-called supercomplexes. This model is supported by several experimental

findings. First of all, these supercomplexes were resolved by native polyacrylamide gel techniques (Eubel *et al*., 2003; Schägger *et al*., 2000). Other support for the model was shown by in-gel activity measurements in native polyacrylamide gels (Eubel *et al*., 2004; Schägger *et al*., 2000). Furthermore, electron microscopy studies revealed defined interactions of OXHPOS complexes within the OXPHOS supercomplexes (Bultema *et al*., 2009; Dudkina *et al*., 2005a; Dudkina *et al*., 2005b; Eubel *et al*., 2004; Heinemeyer *et al*., 2007; Minauro-Sanmiguel *et al*., 2005; Schäfer *et al*., 2006; Schägger *et al*., 2000). In addition, flux control experiments confirmed that the respiratory chain operates as one functional unit (Bianchi *et al*., 2004; Boumans *et al*., 1998). The stability of one OXPHOS complex can be affected by one single point mutation introduced in genes encoding another complex. For example, it was shown in mouse and human cell cultures that complex I was only stable in the presence of mutant-free complex III. One single point mutation in complex III affected the stability of complex I unfavourably (Acín-Pérez *et al*., 2004). In a similar way, complex IV was shown to be essential for assembly or stability of complex I in mouse fibroblasts (Diaz *et al*., 2006). Further support was obtained from oxygen uptake assays with isolated potato mitochondria, which correlated the oxygen uptake with the abundance of supercomplexes (Eubel *et al*., 2004). Also the need for cardiolipin for the formation of some OXPHOS supercomplexes (Pfeiffer *et al*., 2003; Zhang *et al*., 2002) is in favour of the solid-state model.

The focus in chapter 2 is mainly one the structure of respiration complexes and supercomplexes. Therefore, a more detailed description of these large protein complexes will follow in the section below.

Respiratorial supercomplexes

Besides the fact that three of the four OXPHOS complexes can associate into specific supercomplexes composed of 1-2 copies of two different complexes, three OXPHOS complexes (I, III and IV) can also form the so called respirasome ($I + III_2 + V_{1-2}$). This supercomplex can autonomously carry out respiration in the presence of ubiquinone and cytochrome *c*. Interestingly, a recent publication on the mouse OXPHOS system reported that by native PAGE techniques a supercomplex containing also complex II was detected (Acín-Pérez *et al*., 2008). It is also believed that respiratory supercomplexes co-exist in the inner membrane with single OXPHOS complexes.

Complex I can form a stable association with the complex III dimer of the respiratory chain (Eubel *et al*., 2003; Schägger *et al*., 2000) and its structure has been studied by electron microscopy and single particle image analysis. This technique will be briefly

discussed in more detail below. In top-view projection maps, the membrane arm of complex I has a slightly curved shape. The dimeric complex III associates to the inside, at the tip of the curved membrane part of complex I in *Arabidopsis* (Dudkina *et al*., 2005a), *Zea mays* (Peters *et al*., 2008) and potato (Bultema *et al*., 2009). The top-view projection maps of this supercomplex in plants are rather similar. The bovine I + III_2 supercomplex shows a similar association between complexes I and III_2 (Dudkina *et al*., 2010a). From this we suggest that the interaction between the two complexes is conserved among many organisms.

A second supercomplex, comprising complex III_2 and one to four copies of complex IV, has been described for potato (Eubel *et al*., 2004), spinach (Krause *et al*., 2005), *Asparagus* (Dudkina *et al*., 2006a) and beef (Schägger *et al*., 2000). Association of these two complexes within the inner mitochondrial membrane depends on the presence of cardiolipin (Pfeiffer *et al*., 2003; Zhang *et al*., 2002). The first detailed projection maps of the mitochondrial III_2 + IV_{1-2} supercomplex were obtained for *Saccharomyces cerevisiae* (Heinemeyer *et al*., 2007). This result was possible because of the high stability of this supercomplex in yeast. Moreover, this organism lacks complex I, which can associate to complex III as described before, leaving complex IV as the only association partner for complex III. Supercomplex particles were imaged under different angles, because they have different orientations on the support films of the EM specimens. This made it possible to closely compare them with high-resolution 3D data from X-ray diffraction studies and to generate a hybrid pseudo-atomic 3D model. The model shows that two monomeric copies of complex IV are attached at opposing sides of dimeric complex III with their convex sides facing the complex III_2. This is in contrast with the side involved in the formation of complex IV dimers as described by X-ray crystallography.

The above mentioned largest OXPHOS supercomplex termed respirasome consists of complex I, dimeric complex III and complex IV (Schägger *et al*., 2000). It appears that in the respirasome Complex III_2 attaches to complex I in a similar way as described for the I + III_2 supercomplex (Dudkina *et al*., 2010a). Because the supercomplex I + III_2 moiety is easily recognizable within the respirasome structure, comparison of the two projection maps reveals a small density at the tip of complex I and next to the complex III_2 as a specific difference. This density could represent a complex IV monomer. The rotational orientation of complex IV could not be determined yet and requires further investigation (Dudkina *et al*., 2010a). Determination of the handedness of complex I + III_2 and the respirasome from the projection maps was quite difficult from the 2D maps only. However, the 3D model of complex I (Radermacher *et al*., 2006) and the 2D cryo-EM projection map of the membrane arm of complex I (Baranova *et al*., 2007) both show a

curved shape for the membrane arm of complex I. This curved feature can be used as a motif for determining the handedness. This indicates that the first model for the structure of the $I+III_2$ supercomplex from *Arabidopsis* (Dudkina *et al*., 2005a) had the wrong handedness. It should be mentioned though, that the curved feature of the membrane arm of complex I was not available at that time. Probably for the same reason, the 3D model proposed for the bovine respirasome (Schäfer *et al*., 2007) was not assigned correctly. Examination of this 3D model viewed from the IMS clearly shows the typical curvature of the membrane part of complex I (Dudkina *et al*., 2010a).

Higher organization of respiratory supercomplexes

In few cases it appears that supercomplexes are further organized into larger structures called strings. The clearest example is the ATP synthase complex (complex V), which assembles into long oligomeric chains (Allen *et al*., 1989; Dudkina *et al*., 2005b; Krause *et al*., 2005; Strauss *et al*., 2008). Recently, a BN-PAGE study showed an odd and even number of associated mammalian ATP synthases within one fragment (Krause *et al*., 2005), while another study revealed only ATP synthase fragments with an even number of ATP synthase complexes (Wittig *et al*., 2008). The latter was also shown for osmotically shocked mitochondria from *Polytomella* (Dudkina *et al*., 2006b), which suggests that higher organisation of ATP synthases is a consequence of multimerization of dimers. To date, 2D maps of dimeric ATP synthase are available for bovine (Minauro-Sanmiguel *et al*., 2005), the colorless green alga *Polytomella* (Dudkina *et al*., 2005b), *S. cerevisiae* (Dudkina *et al*., 2006b) and potato (Bultema *et al*., 2009), all obtained by single-particle electron microscopy. In all organisms two monomers associate via the membrane Fo parts and make an angle, which can be fixed or variable between 35° and 90°, depending on the organism.

Recently cryo-electron tomography (cryo-ET) studies with fragmented bovine and rat mitochondria demonstrated long rows of dimers located at locally curved cristae membranes (Strauss *et al*., 2008). Averaging revealed larger angles between monomers than reported for isolated bovine dimeric ATP synthase (Minauro-Sanmiguel *et al*., 2005). In another study, dimeric ATP synthase was studied in intact *Polytomella* mitochondria embedded in amorphous ice. Cryo-ET experiments showed the presence of oligomeric rows of ATP synthases in mitochondrial cristae (Dudkina *et al*., 2010b). Details of the rows were obtained at a resolution of 5.7 nm by averaging sub-volumes of tomograms; this shows that the oligomeric chain is composed of repeating dimeric units with a spacing of 12 nm. The monomers make contact via the membrane part and the peripheral stalk

and individual dimers resemble very well the projection maps of isolated dimers in negative stain (Dudkina *et al*., 2005b) and amorphous ice (Dudkina *et al*., 2010b). So far, the tomography data convincingly shows the oligomeric state of the ATP synthase complex, but due to the limited resolution interactions between individual subunits could not be revealed yet.

Besides the oligomeric chains of the ATP synthase complex there appears to be another example for the higher level of OXPHOS organization. A row-like organization of OXPHOS complexes I, III and IV into respiratory strings has been proposed, based on biochemical evidence. It was suggested that respirasomes are interconnected with III_2 + IV_4 supercomplexes (Wittig *et al*., 2006). It appears that such transient strings cannot be purified after detergent solubilization, which limits structural studies. However, one can study all its possible breakdown products that result from detergent solubilization. This was done for potato mitochondria where the breakdown products were analyzed with transmission electron microscopy and an extensive analysis of a large heterogeneous single particle dataset (Bultema *et al*., 2009). This resulted in projection maps for at least five different supercomplexes, including I + III_2, III_2 + IV_1, V_2, I + III_2 + IV_1 and I_2 + III_2 supercomplexes in different positions. Based on assignment of the individual complexes within the supercomplexes combined with the different views of these supercomplexes, the structure for the largest unit, the I_2 + III_2 + IV_2 supercomplex, was proposed. This allowed us, in combination with early electron microscopic observations (Allen *et al*., 1989), to propose a consistent model of the respiratory string for potato mitochondria, which is composed of repeating I_2 + III_2 + IV_2 units (Bultema *et al*., 2009). This model modifies and extends the hypothetical scheme presented before (Wittig *et al*., 2006).

Cyanobacteria

In contrast to other photosynthetic bacteria, cyanobacteria perform oxygenic photosynthesis and, together with diatoms and green plants, they are the main producers of oxygen on Earth. Since some of the experiments described in this thesis were carried out with cyanobacteria, we will first describe some important aspects of the organisms.

Ancient microorganisms

Scheele discovered oxygen just over two centuries ago but free oxygen had already been present present on the earth for billions of years. The minimum time for the existence of free oxygen, based on geochemical research, is 2.3 billion years (Bekker *et al*., 2004). Since cyanobacteria are kept responsible for the generation of free oxygen, they should have been present at that time or even longer. And indeed, the study of a 3.4 billion year old sedimentary formation called "Buck Reef Chert", which is located in South Africa, revealed that it was shaped by photosynthesizing organisms that were able to perform carbon fixation (Tice *et al*., 2006). This suggests that these photosynthesizing organisms, or cyanobacteria, are at least 3.4 billion years old and that they developed somewhere between 3.4 billion to 2.3 billion years ago a system capable of oxygenic photosynthesis (Allen *et al*., 2007). Although the time range for the origin of photosynthetic cyanobacteria and free oxygen is rather large, one can conclude that cyanobacteria are one of the oldest groups of species that are still living today.

Morphology

Apart from being very old, cyanobacteria are rather small and usually unicellular. However, some species also grow in colonies, which can be filamentous to branched in form. In some filamentous colonies, cells show the ability to differentiate into several different types. Examples are (i) the normal, vegetative cells that form under favourable growth conditions, (ii) the akinetes that are resting cells resistant to harsh climatological conditions and, (iii) heterocysts that are specialized nitrogen-fixing cells formed during nitrogen starvation. Furthermore, cyanobacteria are widespread and found in a diverse range of environments, including soil, seawater and freshwater habitats.

Photosysthesis and the origin of chloroplast

Within the prokaryotic domain there are four main groups (phyla) of bacteria that contain species performing photosynthesis: (1) Proteobacteria, also known as purple bacteria, (2) Chloroflexi, which were formerly known as green non-sulfur bacteria, (3) Chlorobi, the green sulphur bacteria, and last but not least (4) Cyanophyta, which are more commonly known as cyanobacteria (see Bryant *et al*., 2006 for an overview of prokaryotic photosynthesis).

As said, all four mentioned phyla are able to utilize phototrophic processes where light energy is captured and stored chemically as ATP. However, there is one important feature that sets the Cyanophyta members apart from the other three phyla: cyanobacteria perform oxygenic photosynthesis whereas the other prokaryotes carry out anoxygenic photosynthesis. In other words, cyanobacteria split water and produce oxygen as a kind of "waste" ($2H_2O \rightarrow 4H^+ + O_2 + 4e^-$), whereas the members of the remaining three phyla split other molecules like H_2S to transform light energy into chemical energy.

One trivial name for cyanobacteria is "blue-green algae" since they are photosynthetic and mostly aquatic. This name is convenient for talking about organisms in water that make their own food, but it does not reflect any taxonomic relationship between the cyanobacteria and the real algae. Cyanobacteria are prokaryotes, not eukaryotes, and it is only the chloroplast in eukaryotic algae and plants to which the cyanobacteria are related. Besides being the first producers of free oxygen, there is another great contribution of the cyanobacteria, which is the origin of plants. The chloroplast within plants can be seen as a modified cyanobacterium living within the plant's cells. Sometime between 0.54 and 1.0 billion years ago certain eukaryotic cells engulfed cyanobacteria, which make food for the eukaryote host, in return for a home. This event is known as endosymbiosis, and is also the origin of the eukaryotic mitochondrion.

Membrane Systems in Cyanobacteria

Cyanobacteria have a cell wall similar to Gram-negative bacteria, which comprises a plasma membrane, a peptidoglycan layer, and an outer membrane. What sets cyanobacteria apart from other prokaryotes is an internal membrane system, which is present in virtually all cyanobacteria. This internal membrane system, the so-called thylakoid membrane, separates the cytoplasm from the lumen and contains both photosynthetic and respiratory electron transport chains. These electron transport chains partly utilize the same components in the membrane. The respiratory electron transport chain proteins are also located in the cytoplasmic membrane, which separates the

cytoplasm from the periplasm. However, there are no photosynthetic complexes in the cytoplasmic membrane and therefore, in most cyanobacteria, photosynthetic electron transport occurs solely in thylakoid membranes, whereas respiratory electron flow takes place in both the thylakoid and cytoplasmic membrane systems.

A long-standing controversy concerning the cellular ultrastructure of cyanobacteria has been whether the thylakoid membranes exist inside the cell as separate compartments, or whether they have physical continuity with the cytoplasmic membrane. Several recent structural studies suggest that thylakoids are separated membrane entities in cyanobacteria (Liberton *et al*., 2006; Nevo *et al*., 2007; Schneider *et al*., 2007; van de Meene *et al*., 2006). Possibly, both options, separated membranes or a continuation of the cytoplasmic membrane, could co-exist within cyanobacteria but more research has to be done to end this long-standing controversy. The formation of thylakoid membranes inside chloroplasts is better understood: vesicles that originate from the chloroplast inner membrane are transformed into a completely separated membrane system within the chloroplasts (Vothknecht *et al*., 2001).

Vipp1 and the biogenesis of thylakoid membranes

The biogenesis of thylakoid membranes requires many cellular processes that all have to be organized in a chronologically aligned and controlled way. For many years the genes and proteins that are specifically involved in formation and biogenesis of the thylakoid membranes remained elusive. This changed in 2001 when a protein was found that was important for thylakoid membrane biogenesis in plant chloroplasts and cyanobacteria (Kroll *et al*., 2001; Westphal *et al*., 2001). Upon deletion of the open reading frame *hcf155* in *Arabidopsis thaliana*, the complete absence of thylakoid membranes was observed. This indicated that the encoded protein is involved in thylakoid membrane biogenesis. Furthermore, it was suggested that the protein might be involved in vesicle trafficking between the inner envelope and thylakoid membranes of chloroplasts (Kroll *et al*., 2001). Consequently, the protein was named vesicle-inducing protein in plastids 1 (Vipp1). Depletion of Vipp1 in the cyanobacterium *Synechocystis* PCC 6803 strongly affected the ability of cyanobacterial cells to form thylakoid membranes (Westphal *et al*., 2001) and, consequently, also in cyanobacteria Vipp1 appears to be involved in thylakoid membrane formation. Depletion of Vipp1 in both *Arabidopsis* and *Synechocystis* PCC 6803 affected photosynthetic capability. The photosynthetic dysfunction in *Arabidopsis* is probably caused by dysfunction of the entire photosynthetic electron transfer chain (Aseeva *et al*.,

2007) and *Synechocystis* PCC 6803 showed a significant reduction of active photosystem 1 proteins (Fuhrmann *et al*., 2009).

Vipp1 shows a high degree of sequence similarity to the phage shock protein A (PspA), a protein that stabilizes the cytoplasmic membrane under stress conditions, from *Escherichia coli* and other eubacteria, and Vipp1 can functionally replace PspA in *E. coli* (DeLisa *et al*., 2004). The low-resolution structure of the *E. coli* PspA has been resolved by electron microscopy (EM) and single-particle analysis (Hankamer *et al*., 2004). The homooligomeric PspA organizes into a ring structure with ninefold rotational symmetry and a molecular mass of ~1 MDa. Formation of ring structures has also been suggested for Vipp1 from *A. thaliana* (Aseeva *et al*., 2004). However, the oligomeric organization and structure of both PspA and Vipp1 has been challenged recently. Vipp1 from the green alga *Chlamydomonas reinhardtii* forms very long rod-like structures in vitro (Liu *et al*., 2007), which, according to authors, might function as tracks for the transport of proteins and/or lipids that are required for the biogenesis of thylakoid membranes. Another possible role for these rod-like structures might represent an inert storage form, from which Vipp1 subunits may be rapidly recruited when they are required (Liu *et al*., 2007). Furthermore, it has been suggested that PspA from *E. coli* forms cage-like structures that mimic eukaryotic clathrin coats (Standar *et al*., 2008), rather than ring structures. These contradicting structural data make further studies relevant.

Cyanobacteria: biofuel of the future?

The need for energy is increasing ever since the industrial revolution and it seems that the energy consumption will increase to higher levels year by year. Especially the fast-growing developing economies of Brazil, Russia, India, and China (also known as the "BRIC countries") demand more energy with the increase of the standard of living for their inhabitants, which is about 40 percent of the world's population (~6.8 billion people at the 1st of January 2010, estimated by the United States Census Bureau). This will further increase depletion of the fossil fuel sources and once depleted other sources should provide for our energy. Beside renewable energy (such as sunlight, wind, rain, tidal energy and geothermal heat) biofuels obtained from waste plant/organic material could become very important.

There are three generations of biofuels. The first-generation biofuels are produced with conventional methods from feedstock like wheat or sugarcane. However, there has been a lot of criticism about this way of producing biofuels since it leads to food shortages and price rises. Second-generation biofuels are produced from non-food crops (you can

think of grass) that require development of elaborated technologies that, for instance, mimic the enzymatic digestive processes as observed in ruminants. The third generation comprises a number of different biofuels (bio-diesel, bio-gas and bio-hydrogen), which are produced by microorganisms such as algae and cyanobacteria.

Advantages and disadvantages of biofuels

Biofuel production using microorganisms offers several advantages. First, (i) the efficiency of biomass production per hectare by autotrophic microorganisms (e.g. algae and cyanobacteria) is estimated to reach 5– 30 times that of crop plants (Sheehan *et al*., 1998), (ii) the cultivation of the microorganisms consumes less water than land plants and with the use of marine algae and cyanobacteria usage of freshwater can be completely eliminated, (iii) bioreactors can be located on nonarable land eliminating competition with food production. However, there are also some disadvantages of biofuel production by microorganisms. The concentration of cells within a culture has to be relative low due to the limit of light penetration. This problem can be solved by increasing the surface between cultures and light source in combination with stirring or mixing of these cultures. Such kind of improvements of the bioreactors increases the operational costs. Moreover, the care required for these cultures is more intense than that for growing conventional crops for producing biomass (Li *et al*., 2008).

Bio-fuels

Algae and cyanobacteria require light, carbon dioxide, water and inorganic nutrients (mainly nitrates, phosphates, sulphates, iron and some trace elements) to produce lipids, proteins and carbohydrates that enable them to grow, or regarding this topic, to create new biomass. Some microorganisms can accumulate lipids to a significant portion of their biomass. Solvent extraction of these lipid-rich microorganisms results in an oily mixture, which can be converted to bio-diesel with existing methods (Chisti, 2007). The protein- and carbohydrate-rich residue of the biomass that remains after extraction can be anaerobically processed by fermentative bacteria producing, depending on the chosen conditions and bacteria, bio-gas (Lantz *et al*., 2007) or bio-hydrogen (Rupprecht *et al*., 2006). An advantage of bio-diesel is that its combustion properties are similar to those of diesel (Marchetti *et al*., 2007) and added to diesel they fit with today's engine technologies and distribution infrastructures (Posten *et al*., 2009). Bio-diesel was originally produced from animal fats and plant oils, which competed with food production. However, the cyanobacterial way of bio-diesel production does not compete with food production.

Moreover, the residual waste can be fermented to generate energy that is required for the production of and processing of biomass. Furthermore, the CO_2 generated from combustion of bio-gas or -hydrogen can be recycled directly for the production of biomass. Thus, in an ideal situation, bio-diesel can be produced in a carbon-neutral way, since all the power needed for producing and processing the algae could potentially come from bio-diesel itself and from bio-gas produced by anaerobic digestion of the biomass residue (Li *et al*., 2008).

Bio-hydrogen can also be produced by most cyanobacteria and algae without the processes described above. At anaerobic conditions, hydrogenases get activated and reduce protons to hydrogen (extensively reviewed by Rupprecht *et al*., 2006). This way of bio-hydrogen production by cyanobacteria requires more sophisticated and closed bioreactors (Li *et al*., 2008). Such systems have the advantage that it is easier to prevent infections, to maintain the optimal parameters for cultivation and to reach maximal exposure to light (Hankamer *et al*., 2007). At present, almost all hydrogen is produced from non-renewable resources as natural gas by steam reformation, petroleum refining, coal gasification and nuclear powered water electrolysis. The production of hydrogen by microorganisms is a clean and sustainable way of generating this promising fuel for the future (Hankamer *et al*., 2007). However, whether this fuel will be used to power cars equipped with fuel cells, is still the question. The setting up of a large distribution infrastructure requires big investments. However, hydrogen can also be used in a power plant to generate electricity and for this the infrastructure is already present. Moreover, this way of electricity production doesn't emit CO_2, which is another advantage compared to fossil fuels.

Engineering cyanobacteria

Apart from improving and optimizing the closed bioreactors for cultivation of cyanobacteria, the cyanobacteria themselves can also be altered to produce, for instance, more hydrogen. Fortunately, the genomes of many cyanobacteria are known and freely accessible (http://genome.kazusa.or.jp/cyanobase), which makes it possible for scientists to tune cellular processes such that in the end the cyanobacteria are easier to cultivate or produce more specific biomolecules (lipids, carbohydrates, etc.) or hydrogen. A few examples will be discussed here.

Recently, researchers used several recombinant strategies to make a cyanobacterium lyse at the appropriate time. The nickel inducible lysis system that they took from bacteriophages and incorporated into the genome of *Synechocystis* PCC 6803 proved to be

functional and controllable. This inducible cyanobacterial lysing system eliminates the need for mechanical or chemical cell breakage and thereby lowers the amount of energy needed to process the cyanobacterial biomass (Liu *et al*., 2009). Another study showed that, upon deletion of an assembly gene for a specific hydrogenase, the light-driven H_2 production was increased five times (Cournac *et al*., 2004). In the case of depleting Vipp1 in the cyanobacterium *Synechocystis* PCC 6803, research showed that the thylakoid membrane system was lost (see above; Westphal *et al*., 2001). Moreover, it could well be that Vipp1 is involved in certain biochemical pathways. All this information could be used to enhance the amount of correctly formed thylakoid membranes and thereby increase the efficiency of solar energy conversion. Thus, (genetic) engineering of microorganisms holds considerable promise for the future (Rupprecht *et al*., 2006).

CRISPR/Cas: killing without oxygen

The ongoing war between bacteria and bacteriophages (viruses that infect bacteria) is one that will never end. As a response to viral attack, prokaryotes have developed several defense systems against bacteriophages during millions of years. Generally, these systems interfere with specific stages of bacteriophage infection such as the adsorption step to the membrane, or RNA/DNA injection. On the other hand, bacteriophages found their own way to be smarter than the bacteria. Altogether, both bacteria and bacteriophages are progressively altering their systems to be one step ahead of the other. Here we describe the recently discovered adaptive microbial immune system named CRISPR/Cas. The acronym CRISPR was introduced in 2002 by Jansen and colleagues when they found a novel family of repetitive DNA sequences (Jansen *et al*., 2002). These Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) have now been identified in most of the archaeal and bacterial genomes, with an abundance of ~90% and ~40 %, respectively (Jansen *et al*., 2002; Sorek *et al*., 2008; van der Oost *et al*., 2009).

Mechanism of CRISPR immunity

In order to explain the mechanism of the CRISPR/Cas system it is useful to know the general structural characteristics of a CRISPR/Cas locus. A typical CRISPR/Cas locus, as shown in figure 2, starts with several conserved *CRISPR-associated (cas)* genes followed by a leader sequence. The CRISPRs are located behind the leader sequence and consist of spacer-repeat sequences.

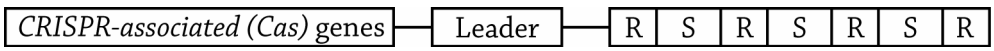


Figure 2. A general representation of the CRISPR/Cas locus as found in prokaryotes. Three different segments can be distinguished. First, the Cas (CRISPR-associated) genes, which interact in some way with RNA and DNA. Second, the leader sequence that might play a role during incorporation of new spacers and possibly can act as a promoter region for transcription of the last part, the CRISPR-units. Each CRIPSR-unit consists of a repeat sequence and a spacer sequence.

The mechanism of the CRIPSR/Cas system (CASS) is not yet fully understood. However, thanks to ongoing research programs that question specific parts of the mechanism, a hypothesis on the general course of action of CASS has been developed. As reviewed (Horvath *et al*., 2010; Karginov *et al*., 2010; van der Oost *et al*., 2009), three particular phases of the mechanism can be distinguished (figure 3): (i) CRISPR immunization, in which foreign genetic elements are recognized by certain Cas proteins,

processed and integrated into the bacterial chromosome as a new repeat-spacer unit at the leader end of the chromosomal CRISPR locus. This provides prokaryotes and archaea immunity against future attacks by bacteriophages with identical genetic elements,

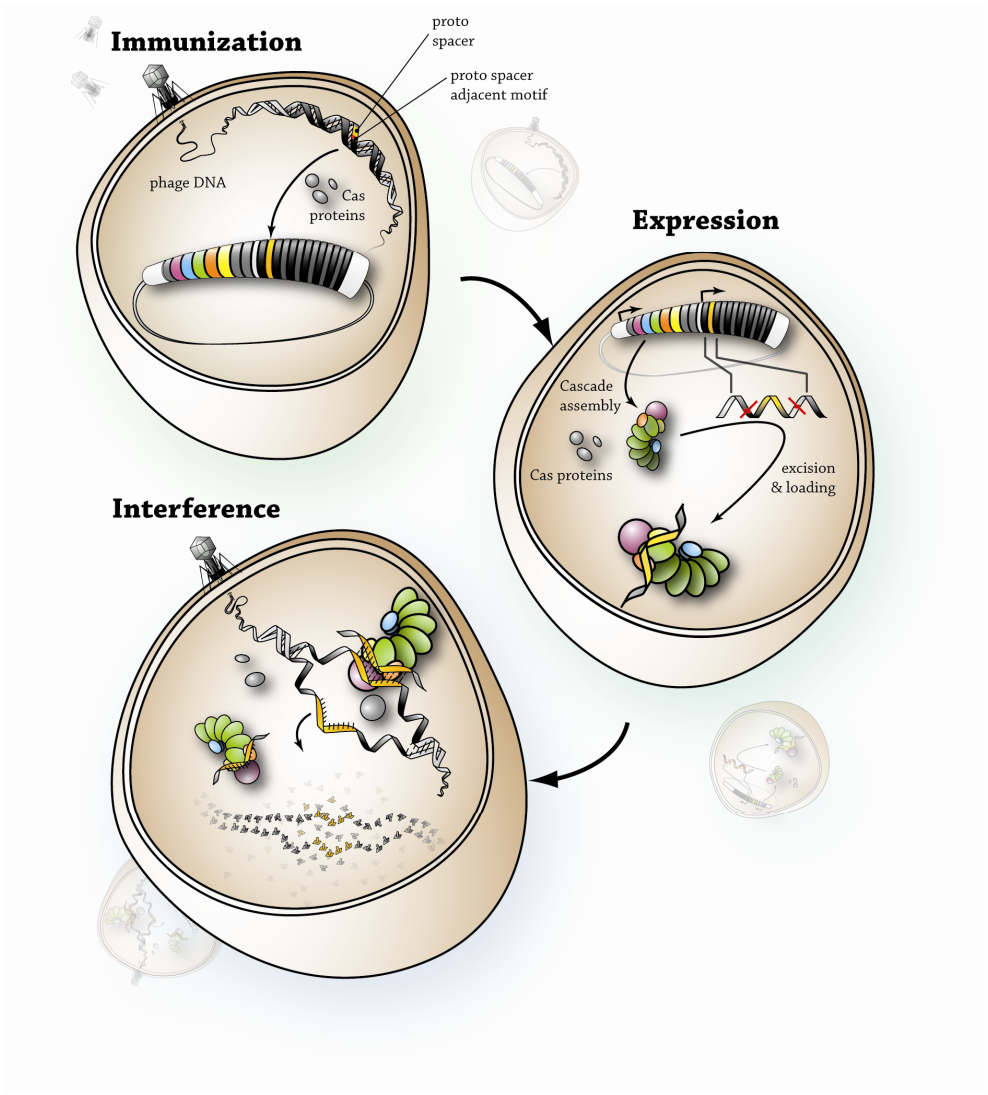


Figure 3. Overview of three distinct phases of the CRISPR/Cas defense system. First, new spacers from injected exogenous dsDNA are recognized by Cas proteins and intergrated at the leader end of the CRISPR locus. Second, the CRIPSR locus is transcribed into pre-cdRNA, which is processed by Cascade into mature crRNAs. During the third phase Cascade uses the bound crRNA as a guide to search for complementary exogenous dsDNA. If there is a match, Cascade and Cas proteins will degrade the exogenous dsDNA preventing proliferation of bacteriophages. Figure adapted from (Karginov *et al.*, 2010).

(ii) CRISPR expression, in which an RNA polymerase binds at a promoter region, which is presumably located in the leader sequence and which transcribes the CRISPR locus into a long single stranded (ss) RNA sequence called pre-crRNA, which in turn is followed by binding to a protein complex called Cascade (CRISPR-associated complex for antiviral defense). This complex processes the pre-crRNAs into smaller RNA sequences consisting of a spacer flanked at both sides by half repeats, also known as crRNAs. (3) These crRNAs act as guide sequences for CRISPR interference whereby complementary foreign genetic elements are blocked such that phage replication isn't successful.

The Cas Protein families

Studies that analyzed the composition of CRISPR/Cas loci and classified the *cas* genes showed that at least four Cas proteins (i.e. Cas1-4) from different families are present in all CRISPR/Cas loci (Haft *et al*., 2005; Jansen *et al*., 2002; Makarova *et al*., 2006). Two other Cas proteins (Cas5 & 6), also from different families, are present in most CRISPR/Cas loci (Haft *et al*., 2005; Makarova *et al*., 2006). This section describes some details known about Cas protein families Cas1, Cas2, Cas3 and Cas6. At present, almost no information is available on the Cas4 and Cas5 families leaving those two unmentioned.

It was suggested that Cas1 could act as a nuclease/intergrase, which might be involved in processing and incorporating new spacers into CRISPR loci (Makarova et al., 2006). Recently, the crystal structure of Cas1 from *P. aeruginosa* was solved (Wiedenheft et al., 2009). Biochemical characterization showed that Cas1 is a metal-dependent DNA-specific endonuclease that cuts double-stranded (ds) DNA fragments into ~80 bp. Since the lengths of these fragments are larger compared to the integrated spacer length (32 nucleotides (nt) for *P. aeruginosa*), the authors expect that Cas1 recognizes or interacts with additional protein(s), nucleic acid(s), and/or nucleic acid modifications (Wiedenheft et al., 2009).

Cas2 proteins are small (less than 100 amino acids) and it was suggested that they have nuclease activity (Makarova et al., 2006). The crystal structure of Cas2 from *Sulfolobus solfataricus* revealed a ferredoxin-like fold (βαββαβ) (Beloglazova *et al*., 2008). A study investigating the function of Cas2 proteins obtained from different organisms showed that they are specific for single-stranded (ss) RNA substrates, preferentially with an uracil-rich region and only in the presence of a metal ion (Beloglazova et al., 2008). Because the examined Cas2 proteins didn't show nuclease activity against ssDNA or dsDNA (Beloglazova *et al*., 2008), it is likely that they are not involved in the first phase (immunization) of the CRISPR/Cas defense system.

Cas3 proteins are big (>100 kDa) and contain a predicted helicase domain together with a small N-terminal HD-domain (Jansen et al., 2002; Makarova et al., 2006). Enzymes with such a HD-domain are either known or are predicted to possess phosphohydrolase activity. They occur in bacteria, archaea and eukaryotes and it has been suggested that they can be involved in nucleic acid metabolism, signal transduction and possibly other functions (Aravind et al., 1998). Biochemical characterization of the CASS from *E. coli* K12 showed that Cas3 is required to protect the cell against phage infection (Brouns et al., 2008).

A biochemical and structural study of Cas6 from *P. furiosus* showed that this protein cuts long pre-crRNA transcripts into small units at one distinct position of a repeat sequence. All resulting RNA fragments are equal in length and each fragment consists of one spacer sequence flanked by a short repeat sequence at the 5′ end and a long 3′ end repeat sequence. Processing of this intermediate RNA fragment into crRNA involves another (Cas) protein which shortens the 3′ end (Carte et al., 2008; Hale et al., 2009) leading to mature prokaryotic silencing RNAs (psiRNA) with two discrete sets of nucleotide lengths (Hale et al., 2008). The structure of Cas6 revealed a duplicated ferredoxin-like fold similar to that of Cas2 (Carte et al., 2008).

Although not much information about the activity of the different Cas protein families is available, it is clear that the mechanism as depicted in figure 3 gives an overall overview. Each phase itself requires presumably several biochemical steps that will be executed by several Cas proteins.

The leader sequence

Leader sequences are AT-rich, vary in length from 20 to 564 basepairs, are non-coding and resemble low structural complexity (Hale *et al*., 2008; Horvath *et al*., 2009; Jansen *et al*., 2002; Lillestøl *et al*., 2006). Although the leader sequences are not conserved in the different species, they can show homology in case several of them are present on the same chromosome (Sorek *et al*., 2008). A study analyzing the leader sequences of lactic bacteria showed that no sequence conservation was observed between identified CRISPR repeat families (Horvath *et al*., 2009). The leader sequence is believed to act as a promoter region for the transcription of repeat-spacer units (Jansen *et al*., 2002; Lillestøl *et al*., 2006). Furthermore, new repeat-spacer units are almost always added between the leader and the previous repeat-spacer unit indicating that the leader may have a function during the acquisition of new spacers (Barrangou *et al*., 2007; Mojica *et al*., 2009; Pourcel *et al*., 2005). Thus, the leader sequences facilitate binding of certain Cas proteins that are

involved in repeat-spacer unit transcription and/or the incorporation of new spacers at the leader end of the chromosomal CRISPR locus.

The repeat and spacer sequences

A typical repeat sequence can be defined as the most frequently occurring sequence within a CRISPR locus. Normally, the repeats within one CRISPR locus are identical but variation among them may occur due to single nucleotide polymorphisms (Horvath *et al*., 2008). Normally, the length of the repeats varies between 21 to 47 bp (Godde *et al*., 2006; Grissa *et al*., 2007; Horvath *et al*., 2010). Each repeat contains a short palindrome of 5 to 7 bp that can form a RNA stem-loop secondary structure (Kunin *et al*., 2007). Beside this structural characteristic, many repeats have a conserved GAAA(C/G) 3′ end which might serve as a recognition site for Cas-proteins (Kunin *et al*., 2007; Sorek *et al*., 2008). A typical spacer is flanked by two repeats and its sequence varies from 20 up to 72 bp (Grissa *et al*., 2007; Horvath *et al*., 2010). In contrast with the sequence homology of the repeats, the spacer sequences are very different and share no homology whatsoever. Studies investigating sequence similarities between CRISPR spacers and extrachromosomal elements showed that the spacers match to phage, prophage, plasmid or chromosomal sequences (Barrangou *et al*., 2007; Bolotin *et al*., 2005; Horvath *et al*., 2008; Horvath *et al*., 2009; Mojica *et al*., 2005; Sorek *et al*., 2008). Spacers and repeats together form a unit and once processed and bound by Cascade, the ss-crRNA (sequence composition: half repeat - spacer - half repeat) acts as a guide sequence for CRISPR interference, whereby complementary foreign genetic elements are blocked such that phage replication isn't successful.

The CRISPR/Cas system of *E. coli* K12

The CRISPR/Cas system from *Escherichia coli* K12 consists of eight *cas* genes that all are located at the 5′ end of the CRISPR locus. The eight genes are *cas3, casABCDE, cas1* and *cas2*. Brouns *et al.* demonstrated that different Cas proteins form a stable complex, which consists of CasA, CasB, CasC, CasD and CasE with one small single-stranded CRISPR-RNA (crRNA) fragment of 61 nucleotides (nt) bound. The complex was denoted as Cascade, CRISPR-associated complex for antiviral defense. The stoichiometry of the proteins within Cascade (determined by native mass spectrometry, chapter 5 of this thesis) was shown to be 1:2:6:1:1 (for A:B:C:D:E).

Binding studies showed that Cascade is guided by the crRNA to dsDNA sequences without the need for supplementary co-factors co-factors such as ATP. In addition to

sequence-specific DNA recognition, Cascade also interacts non-specifically with DNA. Cascade thus appears to interact with DNA in two different modes, non-specific and sequence-specific. It was suggested that the non-specific affinity of Cascade for DNA enables Cascade to associate fast enough with incoming foreign DNA to neutralize a rapidly proliferating phage or conjugative plasmid in a cell (Chapter 5, this thesis). Results from other studies also suggest that crRNAs directly target invader DNA (Brouns *et al*., 2008; Marraffini *et al*., 2008) indicating that the primary target of Cascade for interaction is dsDNA (Chapter 5, this thesis).

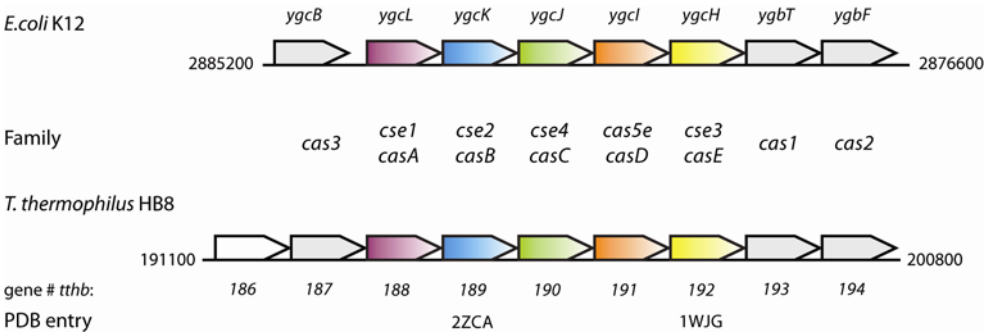


Figure 4. Schematic overview of *cas* genes from *E. coli* K12 and *T. thermophilus* HB8. Genes are drawn as block arrows (not to scale). The *E. coli* K12 primary gene names are given above the block arrows, those for *T. thermophilus* HB8 directly below each block arrow. The Cas protein family nomenclature is as described in (Brouns *et al*., 2008; Haft *et al*., 2005; Houwing *et al*., 2007). The PDB entry of published protein structures is given below the corresponding *T. thermophilus HB8* genes. The proteins that are part of Cascade as found in *E. coli* K12 are coloured purple (CasA), blue (CasB), green (CasC), orange (CasD) and yellow (CasE). The corresponding *T. thermophilus* genes were identified based on sequence alignment with the *E. coli* genes. The grey coloured block arrows represent the proteins Cas1, Cas2 and Cas3. Gene *tthb186,* which is located directly before *tthb187* (*cas3*), was not identified as a Cas protein. The gene maps were obtained from www.genome.jp for both prokaryotes.

The CRISPR locus of *E.coli* K12 is homologous to the *Thermus thermophilus* HB8 CRISPR locus except for an additional gene located adjacent to the *cas3* gene (figure 4). This uncharacterized gene *tthb186* could be a putative transcriptional regulator. The first ~80 amino acids of TTHB186 show sequence homology to the TRAM domain of the crystallized protein RumA (ribosomal RNA 5-methyluridine methyltransferase from *E. coli*) (Lee *et al*., 2004). The TRAM domain, a five-stranded antiparallel β-barrel, is present in several other proteins associated with the translation machinery, and it is predicted to be an RNA binding domain (Lee *et al*., 2004). On the other hand, TTHB186 is characterized as a COG2378 protein family member that does not belong to the CRISPR associated protein family, according to the *cas* gene classifications available today (Haft *et al*., 2005; Makarova *et al*., 2006).

At present, two *T. thermophilus* Cas proteins structures are available, but none of the *E. coli* Cas proteins. On the other hand, a low-resolution structure of *E. coli*'s Cascade is available but structural information of Cascade from *T. thermophilus* is lacking. However, since *E. coli* and *T. thermophilus* have the same CRISPR locus organization (CASS subtype Cse), combining structural and biochemical data could give more insights into the Cse subtype.

CRISPR-based practical applications

To keep things simple, bacteria can either be good or bad for humans. Good bacteria are used for food production as well as non-food applications (e.g. diary products and insulin, respectively). The applied cultures have to stay bacteriophage free; infection would lead to death of all bacteria and high costs due to the loss of the culture and products. Adjusting the CASS of these bacteria such that they become immune for (most common) bacteriophages would result in lower production costs and higher production rates. The "bad" or pathogenic bacteria make people sick. Because of antibiotic abuse, drug-resistant harmful bacteria become more and more problematic. However, CASS may provide one way to defeat the drug-resistant bacteria. Instead of tuning the CASS, the bacteriophages could be engineered in such a way that the CASS from drug resistant bacteria doesn't recognize injected genetic elements anymore, which eventually could lead to their death. To establish any CRISPR-based application, profound and multi-disciplinary fundamental research programs are required.

Remaining questions

Although the knowledge about the CRISPR/Cas systems is increasing rapidly, many questions still remain unanswered. For instance, what are the expression levels? Is the system expressed all the time at a certain level or is it induced only by an attack of bacteriophages? If so, what is the signal and how and at which stage is its transduction regulated?

It is clear that new spacers are incorporated adjacent to the leader sequence, which gives CRISPR/Cas systems their inheritable and adaptive properties. However, the number of repeat-spacer units per CRISPR locus has to be regulated otherwise the locus could expand to detrimental dimensions. Spacer removal has been reported to occur via homologous recombination between two spacers (Deveau *et al.*, 2008; Horvath *et al.*, 2008; Jansen *et al.*, 2002) but the exact underlying mechanism and involved proteins remain yet elusive. Another question related to incorporation of new spacers is the

requirement of an extra repeat sequence located between the leader and the new spacer. Is it recycled when an antiquated spacer is removed?

The challenge for the CRISPR community is to answer all these (and other) questions for all CASS subtypes. Although some CASS components are present in all subtypes, the majority is not. Especially the large variety of the Cas proteins (45 families) makes one think that there is not only one CRISPR mechanism.

Single particle averaging

Electron microscopy projections of particles composed of light elements, such as protein complexes, are very noisy and the raw data cannot be interpreted in detail. To retrieve the signal there is only one way, which is to average large sets of images of the molecules in the same position. By adding hundreds or, if possible, hundreds of thousands of projections the signal to noise improves substantially and trustworthy electron density maps are obtained. The first image analysis methods were developed to process images of periodic systems, such as two-dimensional crystals by Nobel laureate Aaron Klug and his group of colleagues (De Rosier *et al*., 1968). The growth of such crystals is laborious and unpredictable and as an alternative, single particle averaging has been developed (Frank, 2002; van Heel *et al*., 2000). This method does not depend on crystals and can deal with the projections of randomly oriented single molecules. This is a great advantage over methods that need protein crystals. The single particle method has become increasingly popular because it also turned out that high resolution is within reach. Single particle EM has been used in all chapters of this thesis to determine the structure of various macromolecules and the method is described below.

Basically, the method of single particle analysis consists of only a few crucial steps. If projections result from one type of orientation on the support film, averaging is possible when the projections are brought into equivalent positions by computing rotational and translational shifts. After alignment, particles can be summed and by increasing the number of summed projections the noise is gradually fading away. However, macromolecules usually exhibit a full range of orientations on a carbon support film, which means that more elaborated data processing steps are necessary. Treatment of the data set with additional steps like multivariate statistical analysis together with automated classification (see Frank, 2002; van Heel *et al*., 2000) can separate different angular projections prior to the averaging step. As a result of the statistical analysis and classification only those images that are most similar are grouped together. The results of the classifications are “classes” with homogeneous projections. In many cases it turned out that iterative processing of the projections within each class gradually improves the resolution of the final sums. The method is applicable for ice-embedded, unstained specimens, as well as the classical negatively stained particles. Ice-embedded single particles are preferred above air-dried specimens, because they allow higher resolution upon averaging. Very recently, the resolution of single particle averaging was pushed up to

3.3 Å for a virus protein of the aquareovirus ISVP, which allowed not only precise fitting of the polypeptide chain, but also to propose a priming mechanism for cell entry (Zhang *et al*., 2010). Single particle averaging applied to viruses takes advantage of the high-symmetry of the analyzed structures. Atomic resolution of a non-periodic object such as the ribosome has not yet been achieved.

Negatively stained particles are still attractive if the samples cannot be purified to homogeneity or if the protein assembles into multiple types of structures, as will be discussed in chapter 3. In contrast to the widely accepted current belief that negatively staining does not allow to obtain a resolution below 20–25 Å, results on catalase crystals showed that negative staining preserves structural information at the high-resolution range of 4.0 Å (Massover *et al*., 2001). On the other hand, it should also be stated that on the same catalase crystals a better resolution of 2.8 Å was obtained when they were imbedded in ice (Boekema *et al*., 2009). In 2D maps or 3D reconstructions a resolution of 8–9 Å on negatively stained protein complexes is possible. Cryo-negative staining structures below 10 Å were obtained from the multiprotein splicing factor SF3b (Golas *et al*., 2003) and GroEL (De Carlo *et al*., 2008). For rigid and well-stained molecules a resolution of 11 Å can be achieved in 2D maps from only 1,000 summed projections (Boekema *et al*., 2009). But one has to realize that some large protein complexes show a certain structural flexibility under the applied experimental conditions (Yeremenko *et al*., 2004), which limits the achievable resolution. However, even medium resolution (10-15 Å) data of large multi-subunit complexes can be very useful, especially when structures for the individual subunits are known. Fitting of X-ray data into EM data is useful if the EM data are better detailed than 15 Å, because fitting yields hybrid structures, which are about 5 times better than the resolution of the EM structure (Rossmann *et al*., 2001). This means that it is already possible to look at helix-helix interactions with EM 2D maps and 3D volumes slightly better than 15 Å as was done for the mitochondrial III_2-IV_{1-2} supercomplex (Heinemeyer *et al*., 2007). A similar approach was applied on supercomplexes isolated from potato mitochondrial inner membranes (Chapter 2, this thesis). A drawback of membrane-proteins is the fact that they only stay monodisperse in solution within a non-ordered detergent layer (Boekema, 1991), which makes projections fuzzy at the circumference.

The use of slow-scan CCD cameras, which can be programmed to record hundreds of images in a semi-automated way, helped tremendously to extract large data sets necessary for detailed studies. In the near future, it is expected that direct electron counters with superior recording qualities will replace the CCD cameras (Faruqi *et al*., 2007) and that

further automation will provide structures within hours after sample insertion in the microscope. In addition, much higher contrast of unstained specimens is possible by application of “novel” phase contrast electron microscopy such as the Zernike phase contrast microscopy (Yamaguchi *et al.*, 2008). The set-up in an electron microscope is similar to the phase contrast light microscope, for which Frits Zernike was awarded the Nobel Prize for physics in 1953. The recent developments have changed the single particle method from laborious (recording on photographic emulsion, scanning the images by densitometers and processing on less sophisticated stand-alone computers) into the direction of full automation of all steps, i.e. from automated particle collection to detailed 3D reconstructions. Implementation in commercial electron microscopes of Zernike phase plates will be a logical next step in improving EM methods.

Chapter 2

Megacomplex organisation of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato

Jelle B. Bultema, Hans-Peter Braun, Egbert J. Boekema and Roman Kouřil

Published in: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* (2009) 1787: 60-67

Abstract

The individual protein complexes of the oxidative phosphorylation system (OXPHOS complexes I to V) specifically interact and form defined supramolecular structures, the so-called "respiratory supercomplexes". Some supercomplexes appear to associate into larger structures, or megacomplexes, such as a string of dimeric ATP synthase (complex V_2). A row-like organization of OXPHOS complexes I, III and IV into respiratory strings has also been proposed. These transient strings cannot be purified after detergent solubilization. Hence the shape and composition of the respiratory string was approached by an extensive structural characterization of all its possible building blocks, which are the supercomplexes. About 400,000 molecular projections of supercomplexes from potato mitochondria were processed by single particle electron microscopy. We obtained two-dimensional projection maps of at least five different supercomplexes, including the supercomplex I + III_2, III_2 + IV_1, V_2, I + III_2 + IV_1 and I_2 + III_2 in different types of position. From these maps the relative position of the individual complexes in the largest unit, the I_2 + III_2 + IV_2 supercomplex, could be determined in a coherent way. The maps also show that the I + III_2 + IV_1 supercomplex, or respirasome, differs from its counterpart in bovine mitochondria. The new structural features allow us to propose a consistent model of the respiratory string, composed of repeating I_2 + III_2 + IV_2 units, which is in agreement with dimensions observed in former freeze-fracture electron microscopy data.

Introduction

Many processes in the cell require energy which can be stored in a chemical way. Adenosine triphosphate (ATP) is such a storage molecule. Hydrolysis of ATP produces besides inorganic phosphate and adenosine diphosphate (ADP) "free energy" which can be used to drive energy-requiring processes. The main eukaryotic cell compartment for ATP production is the mitochondrion, where energy of oxidation is coupled with the formation of ATP. This is performed by ATP synthase which utilizes an electrochemical gradient of protons created by three respiratory protein complexes (NADH dehydrogenase - complex I, dimeric cytochrome c reductase - complex III_2, cytochrome c oxidase - complex IV) across the inner mitochondrial membrane. Proton translocation depends on the transfer of electrons from NADH or $FADH_2$ through complex I, III, IV and succinate dehydrogenase (complex II) to oxygen which is reduced into water. Electron flow between

the complexes is mediated by two mobile electron carriers, ubiquinone and cytochrome *c*, which are located in the inner membrane and the intermembrane space, respectively (Kroll *et al*., 2001; Saraste, 1999). Nowadays, availability of high-resolution structures of complex II, III, and IV (reviewed by Rich, 2003), the hydrophilic arm of complex I (Sazanov *et al*., 2006) and most of the complex V (or ATP synthase) subunits (Abrahams *et al*., 1994; Dickson *et al*., 2006; Gibbons *et al*., 2000; Stock *et al*., 1999) gives more detailed insight into the function of individual complexes in mitochondrial respiration.

A pioneering electron microscopy (EM) study of the freeze-fractured and deep-etched inner mitochondrial membranes from *Paramecium multimicronucleatum* revealed at least two types of repeating structures in the membrane and thereby gave the first indication of an ordered organization of respiratory complexes in the membrane (Allen *et al*., 1989). The first periodic structure, a double row of particles following the full length of the outer curve of helically shaped tubular cristae, was identified as ribbons of dimeric ATP synthase. In another study, rows of ATP synthase particles were visualized by cryo-electron tomography of *Neurospora mitochondria* (Nicastro *et al*., 2000). Recently, a tomographic analysis of the inner mitochondrial membrane from bovine mitochondria showed the structure of these ribbons of dimeric ATP synthase in greater detail (Strauss *et al*., 2008). The second type of non-random arranged complexes, consisting of large 13 × 22 nm particles with a spacing of 26–30 nm, was interpreted as rows of dimeric complex I (Allen *et al*., 1989), but has not been confirmed by other structural investigations.

Presence of such periodic structures in the inner membrane presumes a specific association between single respiratory complexes. Indeed, Schägger and Pfeiffer (Schägger *et al*., 2000) showed that a mild solubilization of the inner mitochondrial membrane from yeast and bovine heart mitochondria followed by Blue-native PAGE (BN-PAGE) led to a separation of high-molecular bands with a mass between ~ 1.2 MDa and ~ 3 MDa, which were identified as specific supercomplexes of individual respiratory complexes. Using a similar biochemical approach, different types of respiratory supercomplexes were subsequently identified in many other organisms (reviewed in (Dudkina *et al*., 2006; Marques *et al*., 2007; Schägger *et al*., 2000; Wittig *et al*., 2006b)). These supercomplexes are formed by two to three different respiratory complexes with up to four copies, with the following composition: $I_2 + III_2$, $I + III_2$, I + IV, $III_2 + IV_{1-2}$ and $I + III_2 + IV_{1-4}$. The latter supercomplex has been named respirasome (Schägger *et al*., 2000). In addition, dimers have been found for complexes I (I_2) and V (V_2). The dimeric forms of mitochondrial complex I and V seem to be a special type of association for this organelle, because these enzymes are monomeric in, for instance, plant chloroplasts and all eubacteria.

The structural characterization of respiratory supercomplexes is a prerequisite for our understanding of their functioning but also for the higher organization of supercomplexes in the inner membrane. Experimental data indicate that formation of supercomplexes allows enhanced electron transfer rates (electron channelling) and increases the stability of individual OXPHOS complexes. The higher organization of supercomplexes seems to be important for the morphology of the inner mitochondrial membrane and an increase of the protein insertion capacity of the inner mitochondrial membrane (reviewed in (Boekema *et al*., 2007)). Single particle EM and image analysis revealed two-dimensional projection maps of plant supercomplex I + III_2 (Dudkina *et al*., 2005a) and (Peters *et al*., 2008), dimeric ATP-synthase (Dudkina *et al*., 2005b) and (Minauro-Sanmiguel *et al*., 2005), yeast supercomplex III_2 + IV_{1-2} (Heinemeyer *et al*., 2007) and bovine I + III_2 + IV supercomplex (Schäfer *et al*., 2006). Recently, the first three-dimensional structure of bovine respirasome was solved by EM and the arrangement of the individual complexes within this supercomplex was suggested (Schäfer *et al*., 2007). The resolution of this 3D map is, however, still too low to precisely position the smallest of the components, complex IV. Structural work on the dimeric complex V indicated that this supercomplex is kinked in its membrane-embedded parts, leading to a strong local curvature of the membrane (Dudkina *et al*., 2005b). This suggests that the dimeric ATP synthase is involved in shaping the inner mitochondrial membrane folding. So far, there is no direct structural evidence for the existence of a supercomplex containing a dimer of complex I, and finally, complex II appears not to be incorporated in any type of supercomplex.

The existence of defined types of supercomplexes is now well established, but since the various oxidative phosphorylation (OXPHOS) complexes differ in abundance they are considered to not completely form part of supermolecular structures. Rather, OXPHOS supercomplexes and singular OXPHOS complexes co-exist within the inner mitochondrial membrane. On the other hand, it is also likely that higher types of organization exist in many organisms. The string-like bovine megacomplex consisting of ATP synthase dimers (Strauss *et al*., 2008), discussed above, is one specific structure. Based on biochemical data, another megacomplex has been proposed, which was called the respiratory string. It was suggested, that I + III_2 + IV_4 supercomplexes are interconnected with the III_2 + IV_4 supercomplex (Wittig *et al*., 2006b). This suggestion was based on the earlier observations in freeze-fractured electron microscopy specimens (Allen *et al*., 1989). However, considering the known size of the individual complexes, the proposed model cannot explain the observed periodic spacing of 26–30 nm.

The goal of the present study was a structural characterization of the higher organization of respiratory supercomplexes into rows or respiratory strings in the inner membrane.

Potato mitochondria were chosen as a study target. An important reason to use potato mitochondria is the fact that a supercomplex containing two copies of complex I was detected by BN-PAGE (Eubel *et al*., 2003). Because a transient megacomplex such as the hypothetical respiratory string is probably impossible to purify after detergent solubilization we decided to approach the problem by an extensive characterization with single particle electron microscopy of all the possible building blocks of periodic structures. About 400,000 projections of supercomplexes were processed. We obtained two-dimensional projection maps of at least five different supercomplexes, including the supercomplex I + III_2, III_2 + IV_1, V_2, I + III_2 + IV_1 and I_2 + III_2 in different types of position. From these maps the relative position of the individual complexes could be determined in a consistent way. These new structural features allow us to propose a refined model of a respiratory string, which is also in agreement with former freeze-fracture EM data (Allen *et al*., 1989).

Material and Methods

Preparation of mitochondria

Freshly harvested potato tubers (*Solanum tuberosum* var. *cilena*) were purchased from a local farmer and stored at 4 °C. Mitochondria were prepared by homogenizing 200 g of plant material at 4 °C using a Waring blender for 3 × 5 s, filtrated through four layers of muslin. Subsequently, organelles were purified by differential centrifugations and Percoll density gradient centrifugation as outlined previously (Braun *et al*., 1992). Isolated mitochondria were either directly analyzed by gel electrophoresis or stored at minus 80°C.

Purification and characterization of respiratory supercomplexes

The mitochondrial inner membrane protein complexes of potato were solubilized using a buffer with 5% digitonin, 30 mM HEPES, 150 mM K-acetate, 10% glycerol, pH 7.4. The solubilized respiratory supercomplexes were characterized with one-dimensional Blue-native (BN) PAGE (Wittig *et al*., 2006a). Secondly, to establish the subunit composition, gel strips including the resolved protein complexes were placed horizontally onto a second gel. The separation was carried out in the presence of SDS (two-dimensional BN / SDS PAGE). Afterwards, the gels were stained with Coomassie-Blue.

To purify respiratory supercomplexes, solubilized complexes were loaded on a 0.3–1.5 M sucrose gradient containing 15 mM Tris buffer, pH 7.0, 20 mM KCl, 0.2% digitonin and separated by ultracentrifugation (17 hours at 150,000 x g and 4 °C). Afterwards, the

gradients were fractionated and the fractions were examined for protein complexes by one-dimensional BN / SDS PAGE and Coomassie-Blue staining. In some cases, bands from a one-dimensional BN-PAGE gel were cut out and the protein complexes were electro-eluted with elution buffer (25 mM Tricine, 7.5 mM Bis-Tris, pH 7.0 and 0.01% digitonin) at 4 °C (Braun *et al*., 1992). Intactness of electro-eluted protein complexes was checked using another one-dimensional BN-PAGE.

Electron microscopy and single particle analysis

Selected fractions of the sucrose gradient were dialyzed against buffer (15 mM Tris/HCl, pH 7.0, 20 mM KCl and 0.2% digitonin) to decrease the sucrose content and to improve quality of the EM specimens. Electro-eluted protein complexes were directly used for specimen preparation without any further treatment. Negatively stained specimens were prepared with 2% uranyl acetate on glow-discharged carbon-coated copper grids. Electron microscopy was performed on a Philips CM120 electron microscope equipped with a LaB6 tip operating at 120 kV. Images were recorded with a Gatan 4000 SP 4K slow-scan CCD camera at 80,000x magnification at a pixel size (after binning the images) of 0.375 nm at the specimen level with GRACE software for semi-automated specimen selection and data acquisition (Oostergetel *et al*., 1998). Single particle projections were selected from micrographs both by hand and reference-based automated particle selection procedures incorporated into GRIP (GRoningen Image Processing) software. Single particle data sets were analyzed with the GRIP software on a PC-cluster, including multi-reference and no-reference alignments, multivariate statistical analysis, and hierarchical ascendant classification (van Heel *et al*., 2000). Final two-dimensional projection maps were calculated from the best resolved classes, by summing only the best 5-20% of the projections, with the correlation coefficient determined in the alignment step as a criterion.

Results

Purification and characterization of respiratory supercomplexes

Solubilization of isolated potato mitochondria by digitonin and analysis of the solubilized protein complexes by two-dimensional BN / SDS PAGE allowed identification of the respiratory supercomplexes present in potato mitochondria (Fig. 1). In line with previous results (Eubel *et al*., 2003; Eubel *et al*., 2004), significant numbers of supercomplexes

comprising of complexes I, III and IV (III_2+IV, I+III_2, I_2+III_2 and I+III_2+IV) along with dimeric ATP synthase and the five individual respiratory complexes were detected.

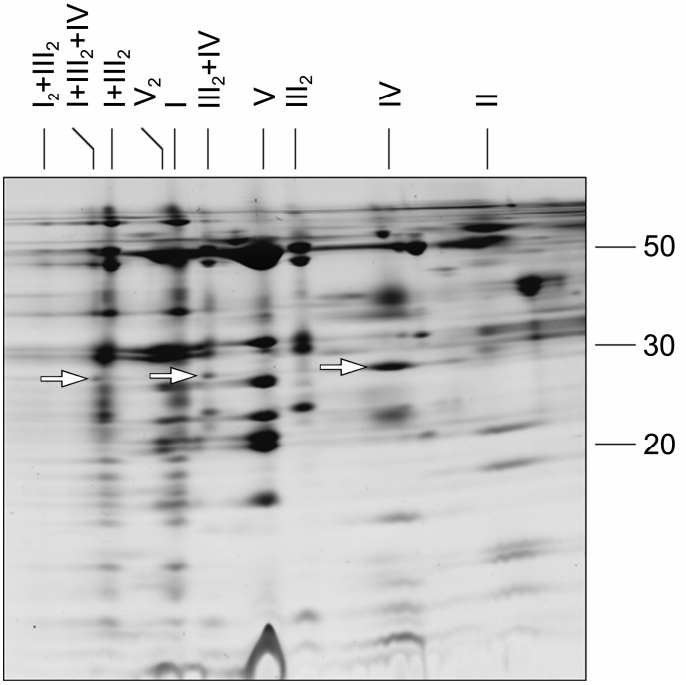


Figure 1. Analyses of protein complexes of potato mitochondria by two-dimensional Blue native / SDS PAGE. Molecular masses of standard proteins are given to the right of the gel (in kDa). Designations: I, II, IV, V: monomeric complexes I, II, IV, V; III2, V2: dimeric complexes III, V; III_2+IV: supercomplex formed of dimeric complex III and monomeric complex IV; I+III_2: supercomplex formed of complex I and dimeric complex III; I_2+III_2: supercomplex formed by dimeric complex III and two copies of monomeric complex I; I+III_2+IV: supercomplex formed of complex I, dimeric complex III and complex IV. The white arrow indicates the Cox II subunit of complex IV, which forms part of (i) complex IV, (ii) of the III_2+IV supercomplex and (iii) of the I+III_2+IV supercomplex.

For single particle EM analysis, solubilized respiratory supercomplexes were purified by sucrose density gradient ultracentrifugation. The gradient was subsequently fractionated and small aliquots of the fractions were analyzed by BN PAGE to monitor their protein complex composition (Fig. 2A). Fractions enriched in specific respiratory supercomplexes were used for single particle EM analysis. If necessary, described purification approach was repeated several times to obtain a fraction with a sufficient concentration of a specific supercomplex suitable for single particle image analysis. Alternatively, EM analysis was performed on electro-eluted supercomplexes from a one-dimensional PAGE strip. To confirm the intactness of the supercomplex after electro-

elution, small aliquots of the electro-eluted supercomplexes were analyzed again with one-dimensional BN PAGE (Fig. 2B).

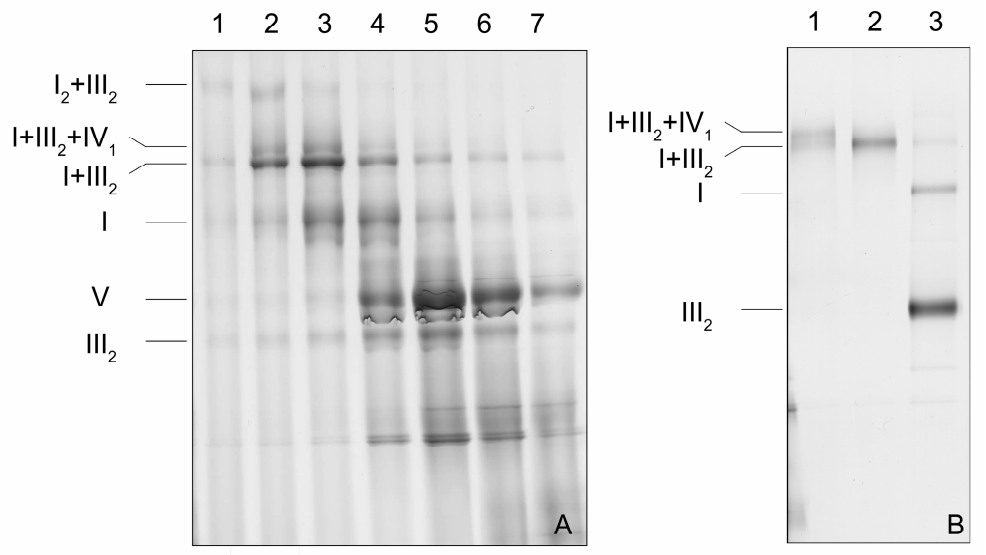


Figure 2. Preparation of mitochondrial OXPHOS supercomplexes from potato for EM analyses. Protein complexes of potato mitochondria were either separated by sucrose gradient ultracentrifugation or by one-dimensional Blue-native PAGE and subsequent electro-elution. (A) Aliquots of fractions of the sucrose gradient were resolved by one-dimensional Blue native PAGE to monitor the protein complex composition of the fractions. (B) Aliquots of electro-eluted $I+III_2+IV$ supercomplex (lane 1), $I+III_2$ supercomplex (lane 2) and complex III_2 (lane 3) were reanalyzed by one-dimensional Blue native PAGE to monitor purity of the fractions. Identities of the resolved protein complexes are indicated to the left of the gels (for designations see Fig. 1).

Structural analysis of complex I

Single particle EM and image analysis of sucrose gradient fraction enriched in content of complex I revealed a top-view and several types of side-view projections of complex I (Fig. 3). Averaged projection maps of the top-view of complex I clearly indicate that the membrane part of the complex I is characteristically curved (Fig. 3A). Side-view projections of complex I were resolved for two different orientations on the carbon film (Fig. 3B-F). More detailed comparison of side-view projections indicates that the main differences are in the structure of the hydrophilic arm (compare Fig. 3C, D vs. B, E, F), although small differences in the membrane part are present as well. As the membrane part of the complex I in Fig. 3C and D appears to be more bent compared to the other side-view projections (Fig. 3B, E, F), different structural features of the hydrophilic arms can reflect a different orientation of complex I on the carbon support film. The reason for the slightly different orientation of complex I can be either in flexibility or opposite conformation of the hydrophilic arm, which can, subsequently, lead to distinct positions

of the complex on the carbon film. The carbonic anhydrase domain, which was found to be tightly associated with the membrane part of complex I from green plants (Dudkina *et al*., 2005a; Peters *et al*., 2008; Sunderhaus *et al*., 2006), can also be recognized in our projection maps. Moreover, the trimeric arrangement of carbonic anhydrase domain was clearly resolved in some projection maps (Fig. 3D, E). Appearance of carbonic anhydrase showing its trimeric structure in the side-view projections evidently indicates its flexible attachment to the membrane part of complex I. The flexibility of the carbonic anhydrase domain is probably located in a flexible connection between the trimeric moiety and the membrane anchor.

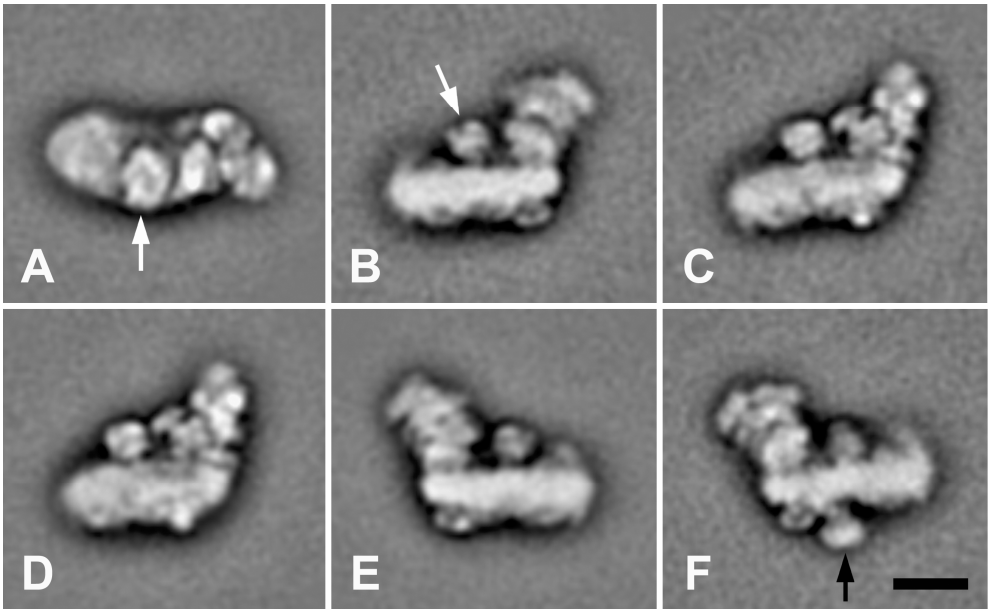


Figure 3. Single particle electron microscopy of the complex I from potato mitochondria. Projection map (A) represents the top-view of the complex I. (B-F) show side-view projections of the complex I with different orientations on the support carbon film, respectively. Carbonic anhydrase domain (at white arrow), which associates to complex I in plant mitochondria, is clearly resolved in all presented projections. An unknown density attached to the membrane part of complex I was revealed in some side-view projections as indicated by the black arrow in (F). Projection maps (A-F) are averaged class sums of 454, 276, 241, 516, 634, and 314 aligned particles, respectively. The scale bar equals to 10 nm.

Interestingly, one class of the side-view projections of complex I reveals a novel density of unknown composition, which is attached to the membrane part of the complex I, approximately on the opposite side of the carbonic anhydrase (Fig. 3F). Estimated from its size, which is slightly smaller than the carbonic anhydrase domain of 75 kDa, it could have a mass between 40 and 60 kDa.

Structural analysis of the I+III$_2$ supercomplex

The projection maps in Fig. 4 show the results of single particle EM and image analysis of sucrose gradient fractions enriched in the potato I+III$_2$ supercomplex.

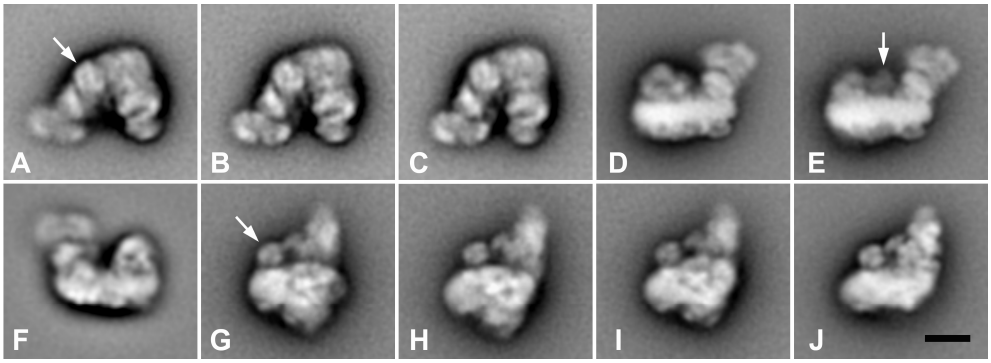


Figure 4. Single particle electron microscopy of the I+III$_2$ supercomplex from potato mitochondria. Projection maps (A-C) represents top-view projections of the I+III$_2$ supercomplex in a more open (A), intermediate (B) and close conformation (C). (D-F) depict side-view projections of the I+III$_2$ supercomplex with different orientations on the support carbon film. Projections (G-I) show tilted views of the I+III$_2$ supercomplex. Projection map (J) represents complex I with a similar orientation on the support carbon film like in case of tilted views of the I+III$_2$ supercomplex shown in (G-I). Projection maps (A-J) are averaged class sums of 1,422, 924, 1,320, 1,024, 1,024, 1,500, 512, 512, 512 and 1,500 aligned particles, respectively. The carbonic anhydrase is indicated by white arrows in (A), (E), and (G). The scale bar equals to 10 nm.

Top-view projections revealed a typical U-shaped feature of the I+III$_2$ supercomplex (including the density of the carbonic anhydrase domain) which have been observed before in other plant species (Fig. 4A-C) (Dudkina *et al*., 2005a; Peters *et al*., 2008; Sunderhaus *et al*., 2006). Detailed inspection of the resolved classes shows that some represent the I+III$_2$ supercomplex in a more opened conformation (Fig. 4A) compared to others (Fig. 4B and 4C). In addition, the carbonic anhydrase domain, attached to the center of the membrane arm of complex I, is distinguishable in some top-view projections as the most stain-excluding density (see arrow, Fig. 4A). Figures 4D-F represent two different side-view projections of the I+III$_2$ supercomplex, which clearly depict typical features of the hydrophilic arm of complex I and the dimeric complex III. The carbonic anhydrase domain can be distinguished only under a certain angular view (arrow, Fig. 4E). In the other two side views, the angular position is such that the density of the carbonic anhydrase is less recognizable as it merges with densities of either the membrane-protruding part of complex III$_2$ (Fig. 4D) or the hydrophilic arm of complex I (Fig. 4F). Besides top- and side-view projections, several classes with tilted supercomplex I+III$_2$ views were found (Fig. 4G-I). Assignment of complex I in the tilted view projections was

facilitated by localization of the trimeric density of the carbonic anhydrase domain, which is composed of three subunits, on the hydrophilic arm of the complex I (Fig. 4G-I). They correspond to the projection densities of a single complex I at a similar angular view (Fig. 4J). The typical feet-like features and characteristic shape of membrane-protruding part of complex III_2 are recognizable in the right-bottom part of the projection maps (Fig. 4G-I).

Structural analysis of I_2+III_2 and I+III_2+IV supercomplexes

Biochemical analysis indicated the presence of large supercomplexes in some sucrose gradient fractions, like the I_2+III_2 and I+III_2+IV supercomplexes (Fig. 2A). Single particle EM and image analysis of these fractions revealed, indeed, the presence of large particles, although their abundance was rather low, with about 1-2 copies per EM micrograph.

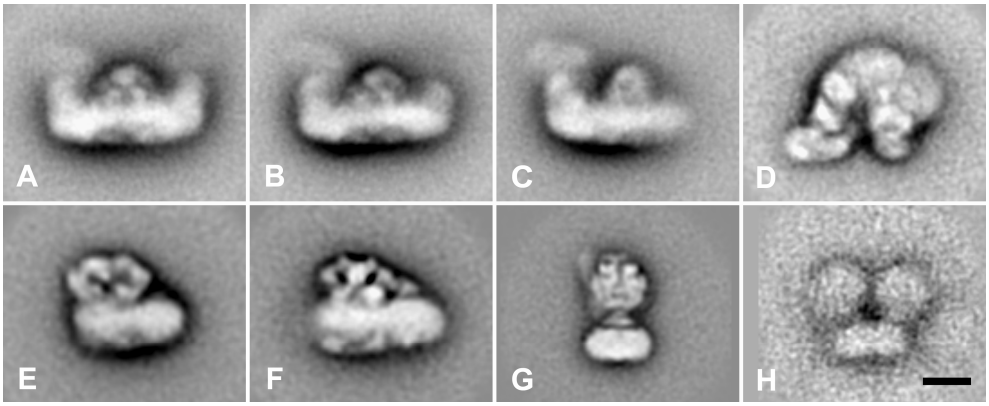


Figure 5. Single particle electron microscopy of the I_2+III_2, I_2+III+IV, III_2+IV, V, and V2 supercomplexes from potato mitochondria. Projection map (A) represents the side-view of the I2+III2 supercomplex whereas (B) shows the same supercomplex lacking parts of the NADH-oxidazing domain in one copy of complex I. Projection map (C) represents a side-view of the I+III2+IV supercomplex, (D) represents the top-view the I+III2+IV supercomplex. (E) represents a side-view of supercomplex III2+IV1. Projection maps (F) represents, for comparison, potato's counterpart of supercomplex III2+IV1 found in yeast (Heinemeyer *et al*., 2007). (G) and (H) represent side-view projections of either monomeric or dimeric ATP-synthase, respectively. Projection maps (A-H) are averaged class sums of 512, 512, 512, 238, 1,024, 768, 642, and 50 aligned particles, respectively. The scale bar equals to 10 nm.

Fig. 5A shows a class average of one type of such a large particle, which was assigned to a side-view projection of the I_2+III_2 supercomplex. The center of the projection map can be unambiguously assigned to the dimer of complex III, mainly due to its typical feature of the membrane-protruding part. Complex III_2 is symmetrically flanked by two copies of complex I, whose hydrophilic arms are recognizable. Image analysis revealed that in some cases the complex I at the right-side of complex III_2 is incomplete in the hydrophilic arm

(Fig. 5B). Importantly, the incompleteness of the I_2+III_2 supercomplex does not lead to a different orientation of the particle on the carbon support film and both the complete I_2+III_2 supercomplex and the fragment reveal the same features in resolved projection maps (compare Figs. 5A and 5B). Fig. 5C shows a class average, which brings to light substantial differences of the complex III_2 compared to its corresponding density in the complete/incomplete I_2+III_2 supercomplex (Fig. 5A, B). Due to the strong similarity of projection map in Fig. 5C with the side-view projection of the I+III_2 supercomplex (4F), we suggest that the projection map in Fig. 5C represents the side-view of the I+III_2 supercomplex with an extra mass attached to the right side of the supercomplex. Furthermore, the similarity indicates that the extra mass has to represent a protein of a smaller size than e.g. the hydrophobic arm of complex I, as there is a minimal influence of the extra mass on the side-view orientation of the I+III_2 supercomplex on the carbon support film. Fig. 5D shows a class average, which represents a well-recognizable top-view projection of the I+III_2 supercomplex with an extra density attached near complex III along the interaction interface between complex I and complex III. Based on the biochemical analysis of sucrose gradient fractions (Fig. 2A) and the size of the extra mass, we suggest that this density represents complex IV and projection maps in Figs. 5C and 5D represent the side- and top-view of the I+III_2+IV supercomplex, respectively. This supercomplex represents the largest individual unit in OXPHOS. It has been found previously in bovine mitochondria and was named the respirasome (Schägger *et al*., 2000).

Structural analysis of the III_2+IV supercomplex and complex V

To prove the existence of a specific association between complex III and complex IV, we performed single particle EM analysis of the III_2+IV supercomplex. Compared to yeast mitochondria, abundance of this supercomplex is rather low, possibly reflecting a weaker association of the two involved OXPHOS complexes. Due to its low abundance in sucrose gradient fractions, we analyzed the III_2+IV supercomplex from an electro-eluted one-dimensional BN-PAGE band. Fig. 5E shows a side-view projection map of the III_2+IV supercomplex, which reveals a specific association between complex III_2 and one copy of complex IV. The overall structure of the side-view projection of the potato supercomplex III_2+IV is similar to the one found in yeast (Heinemeyer *et al*., 2007) (Fig. 5F), but their precise comparison is difficult due to a lower resolution of the potato projection map. In addition to the III_2+IV supercomplex, single particle EM analysis revealed side-view projections of monomeric ATP synthase (complex V) and few copies of dimeric ATP synthase (V_2) in electro-eluted fraction (Fig. 5G and 5H). The structure of monomeric

complex V revealed the F0 and F1 part as well as the stalk. Although the resolution of dimeric ATP synthase is very low, the obtained projection map indicates a small angle between associated monomers.

Discussion

The existence of a higher organization of respiratory complexes into rows or respiratory strings was indicated in intact mitochondria by both freeze-fracture EM studies and electron tomography experiments (Allen *et al*., 1989; Nicastro *et al*., 2000; Strauss *et al*., 2008). The data showing the linear association of dimeric ATP synthase (complex V_2) are convincing because the bulky F1 parts with a diameter of 10 nm protrude from the membrane (Strauss *et al*., 2008), but the structure and/or composition of the respiratory string is not clear. Purified strings might give an answer, but even under mild solubilization conditions the respiratory strings are disrupted and isolated supercomplexes represent just their fragments. In the end, structural characterization of these fragments could allow us to reconstruct the respiratory string. In this study, we performed an extensive structural characterization of respiratory supercomplexes (I_2+III_2, I+III_2+IV_1, I+III_2, III_2+IV_1, V_2) and singular complexes (complex I and V) isolated from one organism where a higher organization into a respiratory string is expected to be present, based on a BN-PAGE survey. Because about 400,000 projections were analyzed, we are quite confident that all the supercomplex structures which are relevant for a higher type of organization into megacomplexes or respiratory strings have been detected. The final projection maps shown here, however, only represent a fraction of the data set. Many other maps, omitted here because of their redundancy, mostly differ slightly in orientation of supercomplexes or in quality of preserved details.

Single particle EM analysis of both sucrose gradient and electro-eluted fractions indicates that a single complex I and the I+III_2 supercomplex were the most abundant particles. Abundance of the I_2+III_2 supercomplex and supercomplexes containing complex IV was substantially lower. Indeed, former biochemical analysis indicates that (i) the I_2+III_2 supercomplex only becomes visible upon prolonged staining of BN gels and (ii) only about 10 % of complex IV is part of the I+III_2+$IV_{1\text{-}4}$ and III_2+$IV_{1\text{-}2}$ supercomplexes, when the rest of complex IV is in the monomeric state under the conditions applied (Eubel *et al*., 2004).

Projection maps of complex I revealed, beside the detailed features of both the trimeric carbonic anhydrase domain and the hydrophilic arm, a characteristic curvature of

the membrane arm of the complex as it is evident from the top view projection (Fig. 3). The same curvature of the membrane arm of complex I was observed in the 3D structure of complex I from *Yarrowia lipolytica* (Radermacher *et al*., 2006) and 2D crystals of the membrane arm of complex I from *Escherichia coli* (Baranova *et al.*, 2007). Based on the curved feature of the membrane arm of complex I we could unambiguously assign the handedness of our 2D projection maps defining the top-view projection as the view from the matrix side. It is worth mentioning that the top-view projection of complex I (Fig. 3A) does not exactly represent the view perpendicular to the membrane plane. It can be derived from the stain distribution along the membrane arm of complex I that the viewing angle in this projection is about 10° off from being perpendicular. In accordance with previous results, side-view projections of complex I revealed the membrane protruding densities at both ends of complex I membrane arm (Fig. 3B-F) (Radermacher *et al*., 2006; Sunderhaus *et al*., 2006). Interestingly, a novel density was resolved on the intermembrane site of the complex I (Fig. 3F). The extra domain is attached to the complex I membrane arm, approximately at the opposite side of the carbonic anhydrase domain. Complex I particles with this extra domain were not detected in any other pant or animal species. At this moment we cannot determine its composition, but given the size it seems represent a plant specific protein which is slightly smaller than trimeric carbonic anhydrase with a mass around 40-60 kDa.

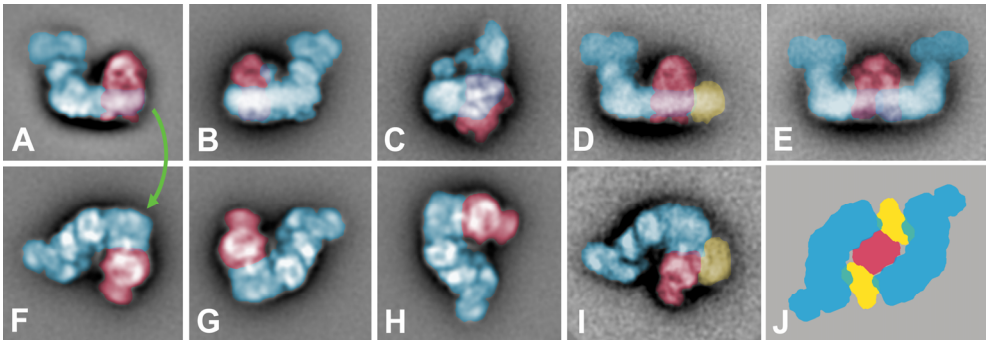


Figure 6. A schematic assignment of different side-views of the I+III_2 (A-C), I+III_2+IV (D), and I_2+III_2 (E) supercomplexes. Green arrow illustrates horizontal flips from side-view projections (A-E) to corresponding top-view projections (F-I). (J) represents a proposed model of a top-view projection of the I_2+III_2+IV_2 supercomplex. Note that the frame of 6I shows half of this structure, but in a slightly tilted position out of the membrane plane. Complex I is depicted in blue, complex III in red, and complex IV in yellow.

Image analysis of the second most abundant supercomplex, the I+III_2 supercomplex, revealed different top-, side- and tilted view projections (Fig. 4). Top-view projections of the I+III_2 supercomplex indicate the same curvature of the complex I membrane arm, as in

the case of singular complex I. In all resolved projections, complex III_2 was found to associate with complex I exclusively at the inner part of the curved membrane arm (Fig. 4A-C, Fig. 6).

An angular difference between the membrane arm of complex I and complex III_2 indicates an apparent flexibility of their association (Fig. 4A-C), which can be important *in vivo*, where the individual complexes associate in the heavily folded mitochondrial inner membrane together with other components. As in the case of the top-view projection of singular complex I, the top-view projection of the $I+III_2$ supercomplex also represents an angular view (~10° tilt out of the membrane plane). This can be evaluated from the amount of deviation from 2-fold symmetry, expected in the projection of dimeric complex III without tilt. Unambiguous assignment of the side- and tilted views of the $I+III_2$ supercomplex could be derived from the top-view projection of the supercomplex (Fig. 6A-C, F-H). Recognition of characteristic features of complex III_2 in both side- and tilted view projections of the $I+III_2$ supercomplex gives us insights into the way dimeric complex III associates to complex I. Fig. 5C and D represent a side- and top-view projection map, which are, except of the presence of a small extra mass, similar to the maps of corresponding views of the $I+III_2$ supercomplex (Fig. 4F, A, respectively). Based on a biochemical characterization of sucrose gradient fractions and dimensions of the extra density, this extra mass was assigned to complex IV and the whole projection map to supercomplex $I+III_2+IV$, the respirasome (Fig. 6 D, I). The top-view projection of the respirasome indicates that the complex IV associates with both complex I and III (Fig. 6I). The resolved projection map of the III_2+IV supercomplex clearly indicates a stable association between these two complexes (Fig. 5E). However, due to the low resolution of the III_2+IV supercomplex it is not possible to determine which site of complex IV is involved in the interaction with complex III_2. Since the overall shape of the supercomplex III_2+IV projection map from potato is rather similar to that of the yeast III_2+IV_1 supercomplex, we assume that in potato complex III_2 and complex IV associate in a similar way like in the yeast supercomplex III_2+IV (Heinemeyer *et al*., 2007).

Our data indicate that the structure of the potato respirasome differs from the 3D model of the bovine respirasome (Schäfer *et al*., 2007). The proposed model for the bovine respirasome suggests that (i) complex III2 associates to complex I from the side, which is opposite to the binding site we found in potato, (ii) complex IV is attached at the end of the membrane arm of complex I, whereas in potato respirasome complex IV associates with complex III_2. So far, neither biochemical data nor single particle EM analysis has given evidence of the existence of the I+IV supercomplex in potato mitochondria, at least under conditions applied in our experiments.

Potato mitochondria are special for I_2+III_2 supercomplexes, although these particles are not very abundant in detergent solubilized membrane fractions. Sucrose gradient fractions where at least trace amounts of the I_2+III_2 supercomplexes were detected by one-dimensional BN-PAGE (Fig. 2A), were subjected to an extensive single particle EM analysis. Fig. 5A shows, for the first time, a side-view projection map of this supercomplex. The middle part of the projection map clearly shows symmetric features, which could be unambiguously assigned to the dimeric complex III. The densities which represent both hydrophilic arms of the two complex I moieties are faint, or even absent on the right position in Fig. 5B. This can be explained by the dissection of the upper part of the hydrophilic arm, which was also observed in complex I from other species (Dudkina *et al*., 2005b). Moreover, the faint appearance is also a matter of stain embedding. Fig. 6J shows a schematic top-view of the I_2+III_2+IV_2 supercomplex. This indicates that in side-view position one of the membrane arms is over 20 nm away from the carbon support film.

The side view map of the I_2+III_2 supercomplex likely also includes two copies of complex IV but because of potentially overlapping positions their presence cannot be ascertained. Together with the top- and side-view maps of the I+III_2+IV supercomplex we can now deduce the relative positions in the full I_2+III_2+IV_2 supercomplex, of which we did not observe substantial numbers top view projections (Fig. 6J) and complete the assignment of the largest possible supercomplex, as deduced from 400,000 potential fragments.

Based on our 2D projection maps of potato supercomplexes we suggest a megacomplex organization in the inner mitochondrial membrane, which was previously named the respiratory string (Fig. 7). We propose that a central role is given to dimeric complex III which is able to bind up to two copies of complex I. Complex IV is supposed to play a role in connection of the I_2+III_2 supercomplexes, and probably also in clasping the flexible interaction between I_2+III_2 supercomplexes. Although EM analysis only revealed a stable association between complex III_2 and one copy of complex IV, binding of two complex IV molecules can be expected, due to the dimeric structure of the complex III. Moreover, biochemical data indicate that the I+III_2 supercomplex can associate with up to 4 copies of complex IV giving the I+III_2+$IV_{1\text{-}4}$ supercomplexes (Eubel *et al*., 2004). The fact that these complexes have not been found in this extensive search may indicate that they easily dissociate upon negative staining procedure used in the EM analysis. In the proposed model, the I_2+III_2+IV_2 supercomplex represents the basic building block (Fig. 7). These supercomplexes form a string by the interaction of two copies of complex IV, whereby each supercomplex contributes one molecule. This organization is based on (i),

the first structure of supercomplex I_2+III_2 which shows that dimeric complex III can bind two copies of complex I, (ii), the structure of supercomplex I+III_2 and (ii), the structure of the III_2+IV supercomplex. Although we were not able to visualize the III_2+IV_2 supercomplex, this supercomplex was present in one-dimensional BN-PAGE gels, making it likely that the supercomplex structure from potato is similar to that found in yeast (Heinemeyer *et al.*, 2007).

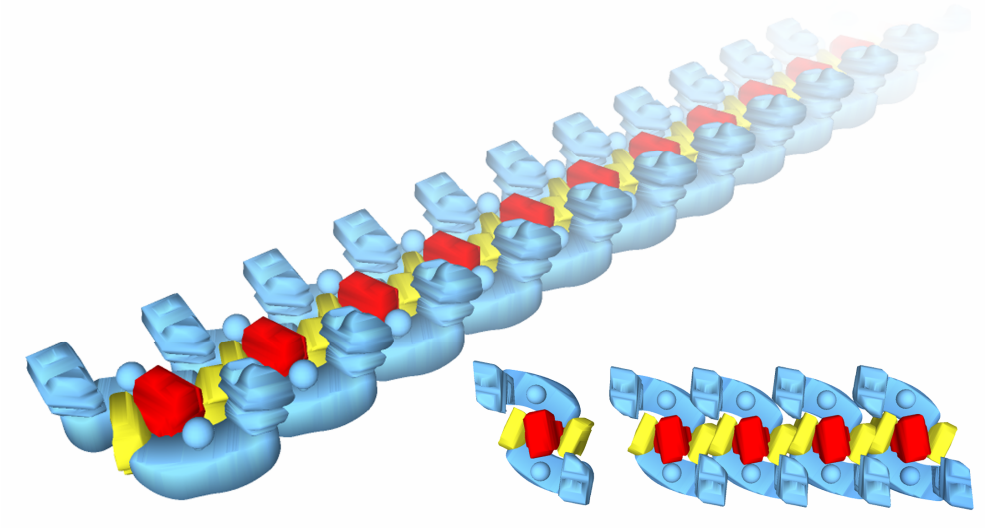


Figure 7. A schematic model on scale of the organization of respiratory chain complexes into respiratory string. The basic unit consists of two copies of complex I (blue), one copy of complex III_2 (red), and two copies of complex IV (yellow). Association of basic units into a string is mediated by complex IV, which interacts with a neighboring complex IV through a dimeric interface found in the X-ray structure (Lee *et al.*, 2001). The model was created using a free version of Google SketchUp software.

The previous model of a string architecture in bovine differs from our model, because it is based on binding of only one copy of complex I to dimeric complex III (Wittig *et al.*, 2006b). Indeed, supercomplexes including more than one copy of complex I were not observed in beef. Further, the ratio of complexes I, III and IV in our model is 1:1:1 and differs from the previous reported ratio of 1:3:6 (Schägger *et al.*, 2001). However, one has to realize that the 1:3:6 ratio reflects the total concentration of each complex in the inner mitochondrial membrane including the free complexes. The I_2+III_2+IV_2 repeating motif in the model proposed for potato measures about 26 nm (going from one centre of complex III_2 to the next) which is comparable to the earlier observed spacing of 26-30 nm by freeze-fractured EM (Allen *et al.*, 1989). Wittig et al. pointed out that this distance equals the length of three repeating units of supercomplexes III_2+IV_2 (see Fig. 3 in Wittig *et al.*,

2006b). This, however, cannot be the case for potato since the length of one III_2+IV_2 supercomplex measured in the configuration IV+III_2+IV equals already 26 nm (Heinemeyer *et al*., 2007). Therefore, it is unlikely that two neighbouring complex III_2's are separated by two dimers of complex IV because addition of two copies of complex IV to the III_2+IV_2 complex would increase the width by about 10 nm to a final 36 nm. Presently it can not be decided whether the string architecture is conserved in eukaryotes or rather differs in different organisms.

In conclusion, the data presented here give the first structural insight in a respiratory string from mitochondria, although in an indirect way, based on the structural characterization of supercomplexes obtained after detergent disruption of the mitochondrial inner membrane. A final proof is in theory possible by electron tomography on ice embedded mitochondria, in a way as performed to visualize the ATP synthase strings (Strauss *et al*., 2008). Tomography, however, has a much lower resolution than single particle analysis. The resolution is around 6-8 nm in the case of large objects such a mitochondrion. In contrast, single particle analysis typically provides information at 2 nm. Hence, present electron tomography can visualize the protruding headpieces of ATP synthase which have a diameter of 10 nm, but it can not reveal the presence of many smaller membrane-embedded components such as complex IV, whose crucial role in the connection of the much larger dimeric complex III and complex I was suggested in our model of respiratory string.

Acknowledgements

We thank Dr. Wilko Keegstra (University of Groningen) for development of a reference-based automated particle selection procedure and Dagmar Lewejohann (Universität Hannover) for expert technical assistance. H.P.B. acknowledges grants of Deutsche Forschungsgemeinschaft (Br 1829-7/3 and Br 1829-8/1) and E.J.B. grants of the Dutch science foundation NWO-CW.

Chapter 3

The Vesicle-inducing Protein 1 from *Synechocystis*. PCC 6803 Organizes into Diverse Higher-Ordered Ring Structures

Eva Fuhrmann, Jelle B. Bultema, Ute Kahmann, Eva Rupprecht,
Egbert J. Boekema, Dirk Schneider

Published in: *Molecular Biology of the Cell* (2009) 20: 4620-4628

Abstract

The vesicle-inducing protein in plastids 1 (Vipp1) was found to be involved in thylakoid membrane formation in chloroplasts and cyanobacteria. In contrast to chloroplasts, it has been suggested that in cyanobacteria the protein is only tightly associated with the cytoplasmic membrane. In the present study we analyze and describe the subcellular localization and the oligomeric organization of Vipp1 from the cyanobacterium *Synechocystis* PCC 6803. Vipp1 forms stable dimers and higher-ordered oligomers in the cytoplasm as well as at both the cytoplasmic and thylakoid membrane. Vipp1 oligomers are organized in ring structures with a variable diameter of 25-33 nm and corresponding calculated molecular masses of approximately 1.6-2.2 MDa. Six different types of rings were found with an unusual 12-17-fold symmetrical conformation. The simultaneous existence of multiple types of rings is very unusual and suggests a special function of Vipp1. Involvement of diverse ring structures in vesicle formation is suggested.

Introduction

Structural and functional studies of the thylakoid (photosynthetic) membranes of both cyanobacteria and chloroplasts revealed many of their exclusive features. However, the biogenesis and formation of thylakoid membranes is incompletely understood today. Chloroplast thylakoid membranes evolve from vesicles originating from the chloroplast inner envelope membrane, and the formation of thylakoids results in the generation of a completely separated internal membrane system (Vothknecht *et al*., 2005; Westphal *et al*., 2001). In contrast, for cyanobacteria it has been a long-standing debate if the internal thylakoid membrane system is physically separated or if it remains connected to the cytoplasmic membrane. Several recent structural studies have, however, indicated that thylakoids are also separated membrane entities in cyanobacteria (Liberton *et al*., 2006; Nevo *et al*., 2007; Schneider *et al*., 2007; van de Meene *et al*., 2006).

For biogenesis of thylakoid membranes many cellular processes, like lipid biosynthesis, membrane formation, protein synthesis in the cytoplasm and/or at a membrane, protein transport, protein translocation, and protein folding have to be organized, chronologically aligned, and controlled. Although for many years genes and proteins specifically involved in formation and biogenesis of thylakoid membranes remained unidentified, in 2001 a protein was described that could be involved in thylakoid membrane biogenesis in plant chloroplasts and cyanobacteria (Kroll *et al*., 2001;

Westphal *et al*., 2001). The observation that deletion of the open reading frame *hcf155* in *Arabidopsis thaliana* resulted in a complete absence of thylakoid membranes indicated that the encoded protein is involved in thylakoid membrane biogenesis. Further analysis has suggested that the protein might be involved in vesicle trafficking between the inner envelope and thylakoid membranes of chloroplasts (Kroll *et al*., 2001). Consequently, the protein was named vesicle-inducing protein in plastids 1 (Vipp1). Depletion of Vipp1 in the cyanobacterium *Synechocystis* PCC 6803 strongly affected the ability of cyanobacterial cells to form thylakoid membranes (Westphal *et al*., 2001) and consequently, also in cyanobacteria Vipp1 appears to be involved in thylakoid membrane formation. A Vipp1 depletion strain of *A. thaliana* is deficient in photosynthesis, and the defects appeared to be caused by dysfunction of the entire photosynthetic electron transfer chain (Aseeva *et al*., 2007). On depletion of Vipp1 in cells of the cyanobacterium *Synechocystis* PCC 6803 a decrease in paired thylakoid membranes has been observed together with a significant reduction of active trimeric photosystem 1 centers (Fuhrmann *et al*., 2009). Although chloroplast Vipp1 was found to be localized in both the thylakoid and inner envelope membrane (Kroll *et al*., 2001; Li *et al*., 1994), the cyanobacterial protein has been suggested to be exclusively bound to the cytoplasmic membrane (Westphal *et al*., 2001).

Vipp1 shows a high degree of sequence similarity to the phage shock protein A (PspA) from *Escherichia coli* and other eubacteria, and - in line with this - Vipp1 can functionally replace PspA in *E. coli* (DeLisa *et al*., 2004). The low-resolution structure of the *E. coli* PspA has been resolved by electron microscopy (EM) and single-particle analysis. The homooligomeric PspA organizes into a ring structure with ninefold rotational symmetry and a molecular mass of ~1 MDa. Formation of ring structures has also been suggested for Vipp1 from *A. thaliana* (Aseeva *et al*., 2004). However, the oligomeric organization and structure of both PspA and Vipp1 has been challenged recently. Vipp1 from the green alga *Chlamydomonas reinhardtii* forms very long rod-like structures in vitro, and a function of these structures in vivo has been suggested (Liu *et al*., 2007). Furthermore, it has been suggested that PspA from *E. coli* forms cage-like structures that mimic eukaryotic clathrin coats (Standar *et al*., 2008), rather than ring structures. These contradicting structural data make further studies relevant.

Here we present a detailed analysis of the oligomeric organization of Vipp1 from the cyanobacterium *Synechocystis* PCC 6803. Vipp1 forms stable dimers and higher-ordered oligomers that are localized in the cytoplasm and at the internal membranes. The higher-ordered oligomeric Vipp1 species have a molecular mass of 1.9 ± 0.3 MDa and are organized in ring structures with a remarkably variable number of monomer copies. They might be involved in vesicle formation by a yet unknown mechanism.

Materials and methods

Heterologous Expression and Purification of Vipp1 from *Synechocystis* PCC 6803

Generation of the plasmid pRSET-SynVipp1 expressing Vipp1 with an N-terminal deca-His-tag was recently described in detail (Fuhrmann *et al*., 2009). For protein expression the expression plasmid was transformed into *E. coli* BL21 (DE3) cells. Cells were grown in 1 L LB medium to an OD600 of ~0.6, and protein expression was induced by addition of 0.5 mM isopropyl-beta-D-thiogalactopyranoside. After 3 h cells were harvested and resuspended in 20 ml 50 mM Na phosphate, pH 7.6, and 5 mM EDTA buffer. Cells were subsequently broken by sonication, and afterward cell debris and inclusion bodies were removed by centrifugation at 10,000 x g for 10 min. The remaining supernatant was directly used to purify Vipp1. The supernatant was loaded onto a Ni-NTA agarose column (Qiagen, Hilden, Germany). The column was washed five times with buffer (50 mM Na phosphate, pH 7.6, 300 mM NaCl, and 40 mM imidazole), and proteins were finally eluted in buffer + 500 mM imidazole. The eluted protein was subsequently dialyzed against 50 mM Na phosphate, pH 7.6.

For determination of the subcellular protein localization in *E. coli*, the supernatant retained after removal of debris and inclusion bodies was centrifuged for 1 h at 100,000 x g, and sedimented membranes were resuspended in buffer and brought to the same volume as the soluble protein fraction. Equal volumes of the fractions were loaded per lane for the SDS-PAGE analysis.

Synechocystis Cell Culture

A glucose tolerant *Synechocystis* PCC 6803 wild-type strain (kind gift of M. Rögner, Bochum, Germany) was grown photoautotrophically in liquid BG11 medium (Rippka *et al*., 1979) at 34 °C under 33 µE/m2 of fluorescent cold white light or under light-activated heterotrophic growth conditions (Anderson *et al*., 1991).

Preparation of cyanobacterial membranes

Synechocystis cells were harvested in the midlog growth phase and pelleted by centrifugation (5,000 x g, 10 min, 4 °C). Cells were resuspended in buffer (20 mM HEPES/KOH, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, and 5 mM $CaCl_2$) and disrupted in a bead beater homogenizer (BioSpec, Bartlesville, OK) using 0.5-mm glass beads. Glass beads, unbroken cells, and cell debris were removed by centrifugation at 5,000 x g for 5 min at 4 °C. After subsequent centrifugation at 100,000 x g for 40 min at 4 °C the

membrane pellet was resuspended in a buffer at a chlorophyll concentration of ~1 mg chlorophyll/ml, and thylakoid membranes were extracted with 1.5% n-dodecyl-β-D-maltoside (DDM) for 1 h at room temperature. Proteins and protein complexes of the soluble fraction as well as extracted membrane proteins were separated on a Biosept S4000 SEC column (Phenomenex, Aschaffenburg, Germany) in buffer (20 mM HEPES/KOH, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, and 0.04% DDM). Individual fractions were analyzed by immunoblot analysis using an anti-Vipp1 antibody (Fuhrmann *et al*., 2009) to identify Vipp1-containing fractions.

For blue native (BN) PAGE analysis freshly prepared *Synechocystis* membranes in buffer (50 mM HEPES/KOH, pH 7.0, 5 mM $MgCl_2$, 25 mM $CaCl_2$, and 10% glycerol) were extracted with 1.5% DDM, and membrane-attached proteins were separated exactly as described detailed before (Dühring *et al*., 2006) . Individual lanes were separated in a second dimension on a 12% SDS-gel (Laemmli, 1970), and Vipp1 was identified by immunoblot analysis using an anti-Vipp1 antibody.

Electron Microscopy

Synechocystis PCC 6803 cells were grown and harvested as described above. A cell pellet obtained from a 10-ml cell suspension was washed three times with EM buffer (50 mM KH_2PO_4/Na_2HPO_4, pH 7.0). The ultrastructural and immunocytochemical (immunogold) investigations were performed as described previously (Gathmann *et al*., 2008).

On purification Vipp1 was dialyzed against 20 mM Tris/HCl, pH 7.5, 75 mM NaSCN, and 50 mM NaCl. Samples of purified Vipp1 were negatively stained with 2% uranyl acetate (UA) or 4% methylamine tungstate (MT) on glow-discharged carbon-coated copper grids. Images were recorded with a Gatan 4K slow-scan charge-coupled device camera (Pleasanton, CA) on a Philips CM12 electron microscope (FEI, Eindhoven, The Netherlands) equipped with a LaB6 tip operated at 120 kV, using Grace software for semi automated specimen selection and data acquisition (Oostergetel *et al*., 1998). The final magnification was 80,000x, with a pixel size (after binning the images) of 3.75 Å at the specimen level.

About 1400 electron micrographs were recorded from a specimen stained with MT. In total 13,943 single-particle projections representing top views were selected and extracted from 1402 electron micrographs, recorded from an MT-stained specimen, by a reference-based automated particle selection procedure incorporated in GRIP (Groningen Image Processing) from these micrographs. From the UA-stained specimen, 600 electron

micrographs were recorded, and 625 single particles of side views were selected and extracted by hand.

Single-particle analysis was performed with the GRIP software package on a PC cluster. The single-particle projections (144 x 144-pixel frame for the top views and 160 x 160-pixel frame for the double-ring side views) were subjected to multireference alignment and reference-free alignment procedure, multivariate statistical analysis, and hierarchical ascendant classification (van Heel *et al*., 2000). Finally, two-dimensional projection maps of Vipp1 top views were calculated from the best resolved classes, which represented ~15% of the whole data. For the final two-dimensional projection maps of the Vipp1 side views, all (80% of the) selected single particles were used to calculate averages from the best resolved classes.

To validate the imposed symmetry, cross-correlation factors (CCFs) between a nonrotated projection map and rotated views of the same projection map were calculated. The rotation angles were such chosen that they correspond to certain symmetry (a rotation of 180° corresponds to twofold symmetry, 120° rotation to threefold symmetry, etc.). Symmetry was identified based on the highest CCF score and the smallest rotation angle because structures with, for example, 12-fold symmetry also show two-, three-, four-, and six-fold symmetry.

Circular Dichroism Measurements

The secondary structure of Vipp1 was determined by circular dichroism (CD) spectroscopy. CD spectra were measured with a Jasco J-810 CD spectrometer (Jasco, Tokyo, Japan) in the 190-350 nm range every 1 nm at a scan speed of 200 nm min^{-1} in a cuvette with a 1-mm path length at a protein concentration of 0.125 mg/ml in 10 mM Na phosphate buffer (pH 7.5). The bandwidth was set to 1 nm.

GALLEX Analyses of the In Vivo Oligomerization in *E. coli*

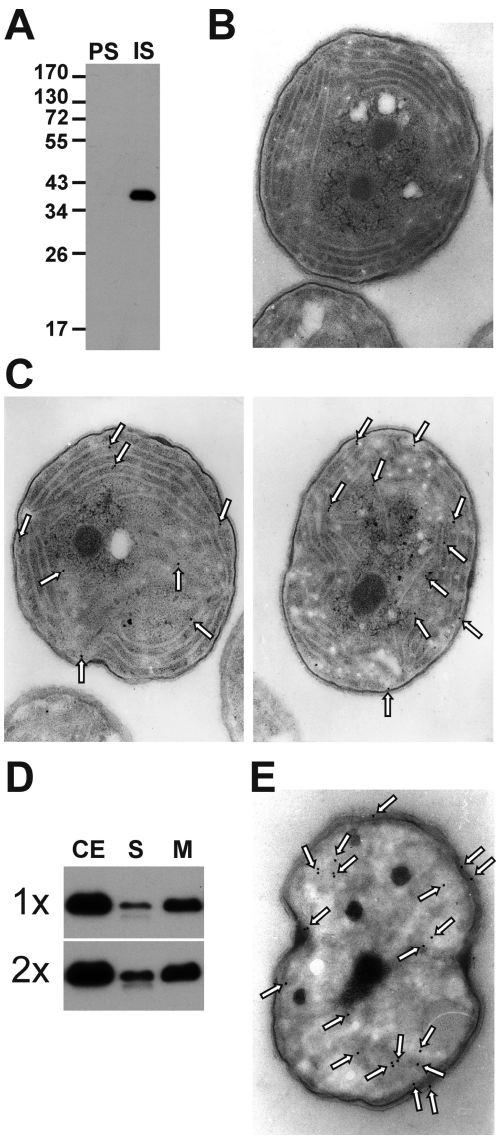
The GALLEX system was used to determine homotypic interactions of Vipp1 in *E. coli*. Therefore, the *vipp1* gene was amplified by PCR from genomic *Synechocystis* DNA using the following primers: vipp3'-GALLEX: 5'-tatgcgagctcgatgggattatttgaccgtttagand-3' and vipp5': 5'-tatgccatatgggattatttgaccgtttagg-3', and the resulting *vipp1* fragment was subsequently ligated to the plasmid pLexA (Schneider *et al*., 2003) after restriction digestion of the PCR fragment and the plasmid with SacI and BamHI. *E. coli* SU101 cells (Dmitrova *et al*., 1998) were transformed with the resulting plasmid pLexA-SynVipp1 as

well as with the control plasmids. Growth of the cells and GALLEX measurements were done exactly as described in detail in (Schneider *et al*., 2003).

Results

Subcellular Localization of Vipp1 in *Synechocystis*

Vipp1 was found to bind to the inner envelope as well as to thylakoid membranes in chloroplasts from *Pisum sativum* and *Arabidopsis thaliana* (Kroll *et al*., 2001; Li *et al*., 1994), whereas it has been proposed that in the cyanobacterium *Synechocystis* PCC 6803 Vipp1 is localized exclusively at the cytoplasmic membrane (Westphal *et al*., 2001). Because chloroplasts and cyanobacteria share a common ancestor this different localization appeared surprising, thus we have reinvestigated the subcellular localization of Vipp1 in *Synechocystis*. To visualize the subcellular localization of Vipp1 in *Synechocystis* cells, we have used EM together with immunogold labeling. As can be seen in Figure 1A the pre-immune serum did not cross-react with any protein from *Synechocystis*, whereas the raised anti-Vipp1 serum showed a clear cross-reactivity with a single protein. Furthermore, when ultrathin sections of *Synechocystis* cells were also probed with the pre-immune serum no gold particle was found on the EM micrographs (Figure 1B). Taken together, these control experiments clearly show that neither the pre-immune serum nor the secondary antibodies cross-react unspecifically with any *Synechocystis* protein different from Vipp1. On the other hand, when ultrathin sections of *Synechocystis* cells were probed with the antibody directed against Vipp1 from *Synechocystis*, the protein was localized at different places within the entire *Synechocystis* cell (Figure 1C). Although most gold particles were found close to membranes, they were not exclusively localized in close proximity to the cytoplasmic membrane but also were close to thylakoid membranes within the cells, which is in good agreement with the described Vipp1 localization in *Pisum* and *Arabidopsis chloroplasts*. About two-third of all gold particles were found in close proximity to membranes, whereas one-third of the gold particles appeared to indicate a localization of (soluble) Vipp1 in the cytoplasm (n = 80). To further test if Vipp1 is exclusively membrane bound in *Synechocystis*, we have examined total cellular *Synechocystis* extract as well as isolated membranes and soluble *Synechocystis* proteins by Western blot analysis. As can be seen in Figure 1D, Vipp1 is mainly associated with membranes but also found in the soluble protein fraction, which indicates a dynamic equilibrium between membrane-bound and soluble Vipp1 in *Synechocystis*. However, in line with earlier reports the majority of Vipp1 was found to be membrane associated.



In the soluble protein fraction a Vipp1 double band was detected by Western analyses, the exact nature of which is elusive. However, a Vipp1 double band has been described in most studies with Vipp1 so far (Aseeva *et al*., 2004; Kroll *et al*., 2001; Liu *et al*., 2005), and this band most likely represents a proteolytic fragment of the full-length Vipp1.

Figure 1. Subcellular localization of Vipp1 in *Synechocystis* PCC 6803. (A) Immunological analysis of *Synechocystis* cellular extracts (5 µg protein) with an anti-Vipp1 antiserum (IS) and the pre-immune serum (PS). Only the anti-Vipp1 serum specifically recognized the Vipp1 protein. Although the calculated molecular mass of Vipp1 is 29 kDa, the apparent mass is ~38 kDa. The molecular masses of the standard proteins are indicated on the left. (B) As observed on Western blots, no protein could be immunostained on electron micrographs with the preimmune serum and the secondary antibody. (C) Electron micrographs of *Synechocystis* cells immunostained with an anti-Vipp1 antiserum (dilution 1:100) followed by treatment with gold conjugated anti-rabbit IgG. The results indicate a preferential localization of Vipp1 in the vicinity of membranes. (D) Total cellular extract of *Synechocystis* (CE) as well as isolated membranes (M) and soluble proteins (S) were analyzed with an anti-Vipp1 antibody. The samples loaded to each lane corresponded to the same amount of cell material. Western blots with two different amounts of the individual fractions are shown to demonstrate that the loaded amounts of cellular extract, soluble proteins, and membranes are in a linear response range for each sample. (E) Electron micrographs of *Synechocystis* cell growth under light-activated heterotrophic growth conditions and immunostaining with an anti-Vipp1 antiserum (dilution 1:100) followed by treatment with gold-conjugated anti-rabbit IgG. The results indicate a localization of Vipp1 in the cytoplasm as well as close to membranes.

To further exclude that the observed fraction of soluble Vipp1 was not simply caused by the release of Vipp1 from membranes during isolation, we further analyzed *Synechocystis* cell growth under light-activated photoheterotrophic growth conditions, where growth does not depend on photosynthesis but solely on glucose (Anderson *et al*., 1991). Under these growth conditions *Synechocystis* cells contain only very few internal thylakoid membranes (Figure 1E). When ultrathin sections of *Synechocystis* cells grown

under such conditions are analyzed with the anti-Vipp1 antibody (Figure 1E), we found a large fraction of Vipp1 not associated with any internal membranes, which further supports the above-mentioned observations with photoautotrophically grown cells.

Taken together, we conclude that, although a large fraction of Vipp1 is bound to membranes, the protein also exists in a soluble form in *Synechocystis*. The membrane-bound fraction is localized at the cytoplasmic as well as at the thylakoid membrane, in agreement with its described subcellular localization in chloroplasts.

Oligomerization of Vipp1 in Synechocystis

To analyze the oligomeric organization of Vipp1 from *Synechocystis* in more detail, we have separated soluble as well as membrane-bound and/or associated protein complexes and subsequently performed BN-PAGE and Western blot analyses.

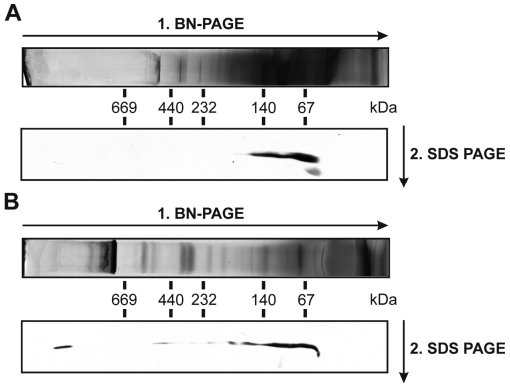


Figure 2. Two-dimensional BN/SDS-PAGE analysis of soluble (A) and membrane-bound (B) proteins and protein complexes from *Synechocystis*. Protein, 1 µg, was separated on a 4.5–16% BN-polyacrylamide gel. Lanes of the BN-PAGE gel (shown on top) were subjected to SDS-PAGE (12% acrylamide) in the second dimension. After Western blotting Vipp1 was identified with an anti-Vipp1 antibody.

As can be seen in Figure 2, after separation of membrane-associated proteins and protein complexes a Vipp1 species with an apparent molecular mass of ~66 kDa was detected as well as a higher-ordered Vipp1 oligomer, which hardly entered the BN gel. These apparent molecular masses correspond to a Vipp1 dimer and a large Vipp1 oligomer with a high molecular mass (>1 MDa), although the exact mass cannot be resolved by BN-PAGE analysis. Interestingly, when soluble proteins from *Synechocystis* were analyzed by BN-PAGE, only the dimeric Vipp1 population was detected. Because this might indicate that the large Vipp1 oligomers form exclusively at membranes, we further analyzed the oligomeric state of Vipp1 by size exclusion chromatography (SEC). This milder technique more likely preserves oligomeric structures. In line with the BN-PAGE analysis a Vipp1

population with a molecular mass of ~66 kDa (dimer) as well as a large Vipp1 oligomer with masses ranging from 1.6 MDa to ~2 MDa were identified (Figure 3). In contrast to the BN-PAGE analysis, the higher-ordered Vipp1 oligomers were found in both the soluble as well as in the membrane protein fraction. These data indicate that Vipp1 forms higher-ordered oligomers in *Synechocystis*, and there may exist a dynamic equilibrium between monomers, dimers, and higher-ordered oligomers.

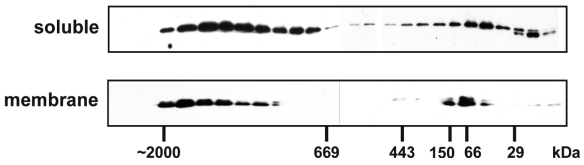


Figure 3. Analysis of soluble and membrane-bound proteins from *Synechocystis* by SEC. Membrane proteins were solubilized with 1.5% DDM and subsequently separated on a Biosept S4000 SEC column (Phenomenex). Identical volumes of individual SEC protein fractions were applied to a 12% SDS gel, and proteins were subsequently analyzed by Western blot analyses. In both the soluble and membrane-bound protein fraction Vipp1 was found as a dimer as well as an oligomer with a molecular mass of ~2 MDa.

However, because the presented in vitro analyses were performed with protein fractions isolated from *Synechocystis* cells, the question arose if Vipp1 also forms oligomers within living cells. To test the oligomerization propensity of Vipp1 in a growing cell culture, we have used the GALLEX system (Schneider *et al*., 2003). In this system a protein of interest is fused to the DNA-binding domain of the *E. coli* LexA protein. Oligomerization of a protein brings (at least) two LexA DNA-binding domains together, and only a dimeric LexA DNA-binding domain can properly bind to a promoter/operator region in the chromosome of an E. coli reporter strain. DNA binding subsequently represses the activity of a reporter gene expressing the β-galactosidase, and, thus, interaction of proteins can be followed by measuring the activity of the reporter. This LexA-based system has initially been used with soluble (Daines *et al*., 2000; Dmitrova *et al*., 1998) but has subsequently been modified for membrane localized proteins (Prodöhl *et al*., 2005; Prodöhl *et al*., 2007; Schneider *et al*., 2004a; Schneider *et al*., 2004b) To evaluate the determined Vipp1 oligomerization propensities properly, we have also analyzed interaction of the transmembrane helix of the human glycophorin A (GpA) wild-type protein as well as of a mutated GpA helix carrying the G83I substitution, which significantly lowers the interaction propensity of the helix (Finger *et al*., 2006; Schneider *et al*., 2003).

The results of the GALLEX measurements are shown in Figure 4. Based on these results, Vipp1 forms stable oligomers in a living cell, and compared with GpA, which

forms a stable helix dimer, Vipp1 has a rather strong oligomerization propensity. For all expressed chimeric proteins we found similar expression levels when tested with an anti-LexA antibody. It has to be noted at this stage, however, that although the GALLEX assay allows a meaningful description of a proteins general oligomerization propensity the measurement does not allow drawing any conclusion about the exact oligomeric state of the protein.

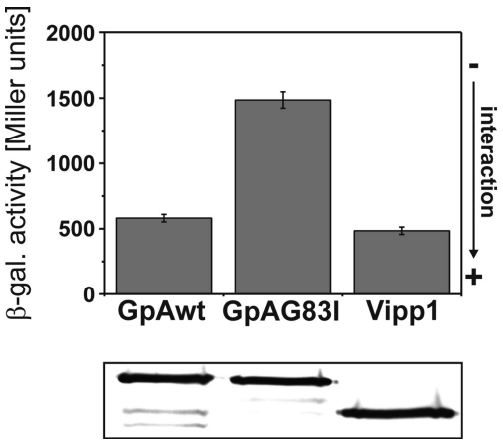


Figure 4. In vivo oligomerization of Vipp1. The propensity of Vipp1 to homooligomerize was measured with the GALLEX-system (Schneider *et al*., 2003). As internal controls, the capacity of the wild-type and G83I-mutated GpA proteins to homo-dimerize was measured. The association capacity of each protein was determined by measuring the β-galactosidase activities of three independent clones. The expression level of the individual fusion proteins was tested by Western blot analysis using an anti-LexA antibody.

Taken together the above presented results strongly suggest that the Synechocystis Vipp1 forms stable oligomers with a mass of ~2 MDa that exist in vivo as nonbound and membrane-bound complexes.

Heterologous Expression of Vipp1

The above described measurements of Vipp1 oligomerization in *E. coli* cells have indicated that Vipp1 can be expressed as a stable oligomer in *E. coli*. Thus, to purify large amounts of the protein for further structural studies, we heterologously expressed the *Synechocystis* Vipp1 in *E. coli* cells. As can be seen in Figure 5, after expression of Vipp1 the protein was found to be mainly associated with membranes. Furthermore, as in *Synechocystis* a small pool of soluble Vipp1 was identified. To subsequently analyze these oligomeric structures in more detail, we have purified the protein from *E. coli* cell extracts (membranes and supernatant) without using any detergent or denaturant, because it has been suggested

that the use of detergents can modify the Vipp1 structure, resulting in non-native structural alterations (Standar *et al*., 2008).

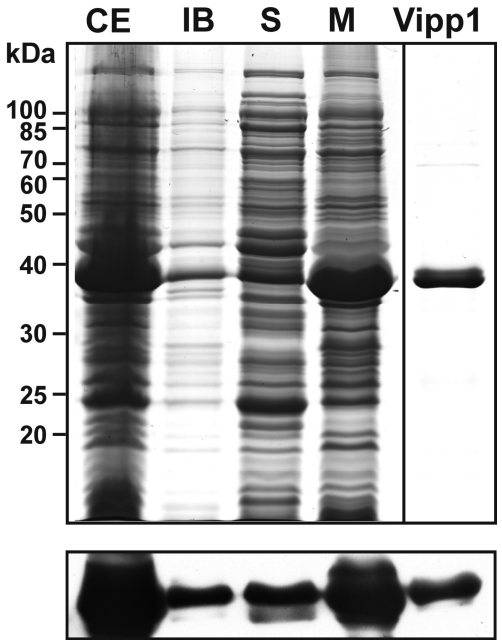


Figure 5. Top, expression of the *Synechocystis* Vipp1 in *E. coli*. Proteins were separated on a 12% SDS gel and stained with Coomassie blue. MW, molecular mass standard; CE, total cell extract of *E. coli* BL21 (DE3) cells harboring the expression plasmid pRSET-SynVipp1; IB, inclusion body fraction; S, soluble proteins; M, membrane proteins. Equal volumes of the individual *E. coli* fractions were loaded per lane. Vipp1 labelled lane, Vipp1 purified by affinity chromatography (1 μg). Bottom, an antibody was directed against Vipp1 cross-reacted with the purified Vipp1 and recognized Vipp1 in the individual *E. coli* fractions.

On purification we have analyzed the secondary structure of Vipp1 by CD spectroscopy. Computational analysis of the Vipp1 amino acid sequence with the program Jpred3 (Cole *et al*., 2008) predicts that ~80% of the protein is α-helical, and many regions are predicted to have a high propensity to form coiled-coil structures (Figure 6A). The CD spectra of the purified *Synechocystis* Vipp1 is shown in Figure 6B and deconvolution of the spectra with the program CDSSTR (Compton *et al*., 1986; Manavalan *et al*., 1987; Sreerama *et al*., 2000; Whitmore *et al*., 2004; Whitmore *et al*., 2008) predicts that 83% of Vipp1 is α-helical, which is in very good agreement with the computational prediction.

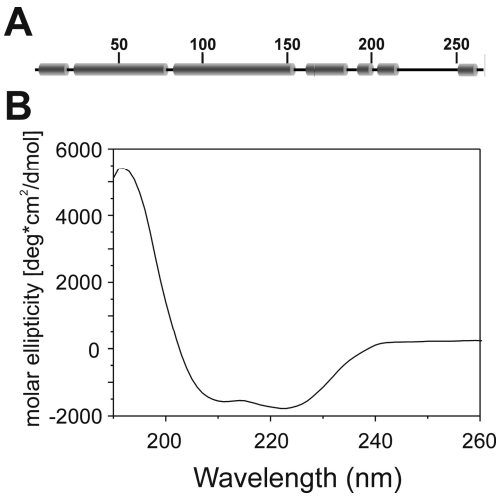


Figure 6. Analysis of the Vipp1 secondary structure. (A) Analysis of the Vipp1 amino acid sequence with the program Jpred3 (Cole *et al*., 2008) predicts that ~80% of Vipp1 is α-helical. The α-helical protein regions are indicated by the bars. (B) CD spectra of the isolated Vipp1 in the range between 190 and 260 nm. Deconvolution of the CD spectra predicts that 83% of Vipp1 is α-helical. Note that the ratio of the ellipticities at 222 and 208 nm is >1. For further details see the text.

Interestingly, the ratio of the molar ellipticity at 222 nm to the molar ellipticity at 208 nm was found to be >1, which is indicative for formation of higher-ordered coiled coil structures (Monera *et al*., 1993). Analyses of purified Vipp1 by BN-PAGE in combination with SEC (Figure 7) showed the same organization of Vipp1 into dimers and higher-ordered oligomers as has been observed before in *Synechocystis* cells (Figures 2 and 3).

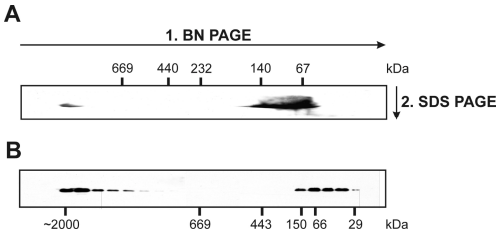


Figure 7. Characterization of the isolated Vipp1. When analyzed by BN-PAGE (A) and SEC (B), the isolated Vipp1 shows an identical pattern of dimers and higher-ordered oligomers as observed in *Synechocystis* cellular extracts (as shown in Figures 2 - 4).

Oligomeric Structure of Vipp1

To further investigate the structure of the Vipp1 oligomers, we have characterized the purified protein fraction by single particle EM analysis. As can be seen in Figure 8, Vipp1 forms large ring structures with different diameters between 25 and 33 nm. The Vipp1

ring structures are composed of distinct units each with a small protein density at the inside and a notable "spike" protruding from the outside (Figure 8A). The presence of the spikes facilitates an accurate determination of the rotational symmetry of the various types of rings. Interestingly, Vipp1 was found to form (at least) six different types of rings with 12–17-fold rotational symmetry (Figure 8, B and C).

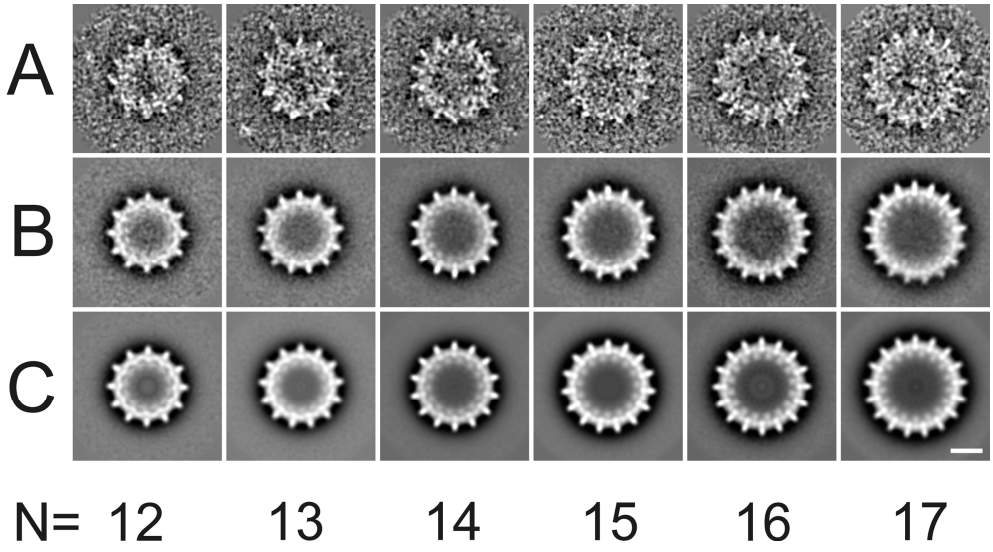


Figure 8. Single-particle image analysis and classification of a set of ~15,000 Vipp1 rings. (A) A gallery of single-particle projections in top-view position. (B) Class-sums of non-symmetrized projection maps of rings with different rotational symmetry. On average, for each projection map ~200–300 projections were summed. (C) Same rings corresponding to the frames of B, with imposed rotational symmetries. Space bar, 10 nm.

Occasionally, two Vipp1 rings become stacked. These double rings were only seen in side-view position (Figure 9D-H). The stacked double rings have a slight variation in diameter, which is between 29 and 34 nm (Figure 9, D–H, horizontal direction). This can be attributed to the above described variation in ring sizes. As expected, the height of the single rings (22 nm, half of the vertical direction) is uniform.

From its shape and dimensions, the mass of a protein complex can be roughly estimated. The 17-fold symmetrical ring has a volume of ~25,000 nm^3, compatible with a mass of 1.5–2.5 MDa. A mass of ~2 MDa was also actually measured. Because the smallest building block of each ring is the Vipp1 monomer with a molecular mass of ~32 kDa and because the ring structures most likely consist of Vipp1 tetramers, rings with 12–17 tetrameric units would result in predicted masses of ~1,500–2,200 kDa. To further assess the nature of the Vipp1 building units shaping the various Vipp1 rings, we have analyzed Vipp1 oligomers by BN-PAGE after mild SDS denaturation of the Vipp1 rings. Addition of

gradually increasing concentrations of SDS to Vipp1 dissolved in 5 mM DDM resulted in the formation of mixed DDM/SDS micelles and gradual disassembly of the Vipp1 oligomers.

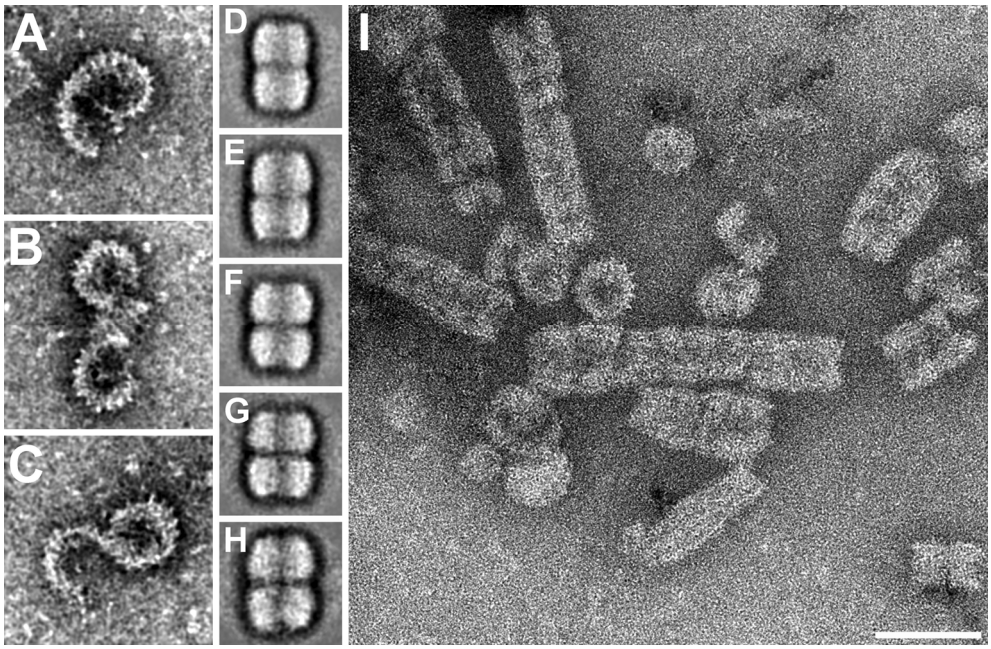


Figure 9. EM nonfrequent types of Vipp1 structures. (A-C) open rings formed by Vipp1. (D-H) Classification of sandwiched Vipp1 rings in side-view position. For each class-sum ~50-100 projections were summed. (I) Overview of rarely occurring rod-like association of Vipp1 rings. In the middle of the frame some individual rings are visible. Space bar for all frames, 50 nm.

At SDS concentrations above 4 mM the large oligomer disappears, whereas dimeric and tetrameric Vipp1 species appear, as well as a small fraction of octameric Vipp1 (Figure 10). At increasing SDS concentrations the higher-ordered oligomers disassemble completely, and only the dimeric Vipp1 remains stable. Thus, in line with the above presented observations Vipp1 forms stable dimers in solution. Therefore, we propose that the repeating unit for a Vipp1 ring is a dimer of Vipp1 dimers. This implies, e.g., for the ring with N = 12 spikes that 24 Vipp1 dimers are involved in formation of the oligomer, with each spike representing a dimer. However, because the spikes facing inside the protein have an electron density different that of the spikes facing to the outside, we suggest that the Vipp1 oligomer is built up by an alternating interaction of Vipp1 dimers, potentially in an antiparallel manner. The side views of the Vipp1 rings indicate that there is no clear separation within distinct domains in the vertical direction (Figure 9, D–H). Because the rings have a height of ~22 nm, this means that more than half of the Vipp1

monomer must exist in an extended α-helical conformation, spanning the entire height of the complex, which is 22 nm. In addition to the observed ring structures, a few string-like open ring structures were found (Figure 9, A–C), and, in addition, a very small part of the entire population of Vipp1 rings was found to form longer rods (Figure 9I).

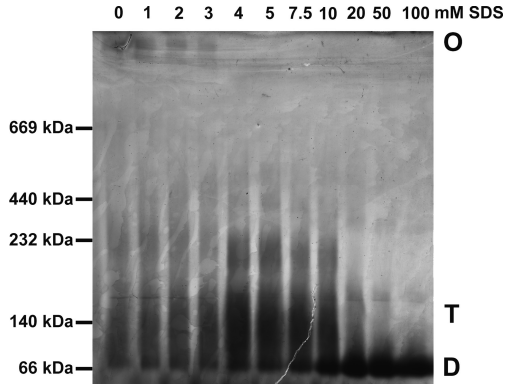


Figure 10. Gradual disassembly of the ~2 MDa Vipp1 oligomer. Vipp1 was dissolved in buffer in the presence of 5 mM DDM, and the indicated amounts of SDS were added to form mixed micelles. After incubation at room temperature for 10 min, protein complexes were separated on a 4.5-16% BN polyacrylamide gel, and the gel was subsequently silver-stained. The positions of the Vipp1 dimer (D), the tetramer (T), and of higher organized oligomers (O) are indicated.

Taken together, the presented results show that Vipp1 from Synechocystis can form ring structures with different diameters but constant heights, and dimeric Vipp1 is the minimal building block involved in formation of the observed higher-ordered oligomeric structures.

Discussion

Subcellular Localization of Vipp1

In the present study we have analyzed the subcellular localization and structural organization of Vipp1 from the cyanobacterium *Synechocystis* PCC 6803. For Vipp1 from *Arabidopsis thaliana*, *Pisum sativum*, and *Synechocystis*, it has been suggested that the proteins are exclusively membrane bound (Kroll *et al*., 2001; Li *et al*., 1994; Westphal *et al*., 2001). Although for the chloroplast Vipp1 a dual localization at both the inner envelope membrane as well as at the thylakoid membrane has been described (Kroll *et al*., 2001), for *Synechocystis* it has been suggested that the protein binds exclusively to the

cytoplasmic membrane and not to thylakoids (Westphal *et al*., 2001). This observation was somewhat unexpected because of the common evolutionary origin of cyanobacteria and chloroplasts and of the proposed function of the protein. If Vipp1 is involved in a vesicular transfer system between the inner membrane systems in chloroplasts and cyanobacteria, one would expect to find the protein localized at both inner membranes as in chloroplasts. In line with the described subcellular localization in chloroplasts the immunogold analysis presented in Figure 1 strongly suggest that cyanobacterial Vipp1 is also attached to both the cytoplasmic and the thylakoid membrane. Thus, in cyanobacteria the protein could be involved in an interlinked communication between the two internal membrane systems as well. However, in contrast to earlier observation a fraction of soluble, nonbound Vipp1 was also identified in *Synechocystis*. Thus, although the majority of Vipp1 appears to be tightly attached to membranes in *Synechocystis*, a soluble protein population is also present, and, as the membrane-bound fraction, soluble Vipp1 also forms higher-ordered oligomers. This observation is in excellent agreement with observations described earlier for PspA from *E. coli*. In *E. coli* PspA exists in a soluble and in a membrane-bound form (Brissette *et al*., 1990), and a dynamic equilibrium between these two stages is suggested to trigger the activity of the protein (Elderkin *et al*., 2002; Hankamer *et al*., 2004). However, it is also possible that cytoplasmic oligomerization of Vipp1 is just a maturation step before membrane binding or even a prerequisite for this.

Structural Organization of Vipp1

Analysis of soluble as well as membrane-bound proteins from *Synechocystis* by BN-PAGE indicates that in *Synechocystis* Vipp1 exists as a stable dimer as well as oligomers with a molecular mass of >>1 MDa. Thus, the protein appears to form larger oligomers than its homologue PspA from *E. coli*. Interestingly, our BN-PAGE analyses indicated that the oligomers were present only in the membrane protein fraction and not in the soluble protein fraction, which indicates that the membrane-bound oligomer is more stable than the soluble one. To further characterize the oligomeric organization of the *Synechocystis* Vipp1, we have subsequently analyzed the protein organization by SEC, and dimeric Vipp1 as well as a Vipp1 population with a significantly higher molecular mass was identified in the soluble as well as in the membrane-bound protein fractions from *Synechocystis*. The molecular mass of the larger oligomer was estimated to be ~2 MDa. Formation of rod-like structures with molecular masses >>2 MDa, as described in (Liu *et*

al., 2007), was not observed, and, thus, the described results strongly suggest that - at least the *Synechocystis* - Vipp1 does not form larger oligomeric assemblies.

After expression of the *Synechocystis* Vipp1 in *E. coli* the monomeric protein strongly interacts and forms higher-ordered oligomers (Figures 5 and 7). Analysis of the purified protein by SEC demonstrated that also the isolated protein forms large oligomeric structures with a molecular mass of ~2 MDa, as observed in *Synechocystis* before. Because an impact of detergent on the structural organization of PspA from *E. coli* has been emphasized (Standar *et al*., 2008), it has to be mentioned that the protein was purified without any detergent. Because the protein can be well purified and, thus, effectively washed off the membranes by this method, this observation may indicate that Vipp1 is not very tightly bound to the internal membranes, at least not in E. coli. This might be caused by the different lipid composition of the *E. coli* membranes compared with the *Synechocystis* membranes. However, as also the homologous protein PspA from *E. coli* was purified by a similar method after homologous expression this protein appears also not to bind tightly to membranes (Standar *et al*., 2008). Weak membrane attachment could thus be a common feature of PspA/Vipp1 proteins.

EM shows that Vipp1 monomers can form various higher-ordered structures, and the BN-PAGE analysis shown in Figure 10 suggests that the these structures are built up by accumulation of multiple dimeric Vipp1 species. Numerous Vipp1 proteins laterally associate to form small strings (Figure 9, A–C), which eventually close to form Vipp1 rings. The strings shown in Figure 9, A–C, could either represent assembly intermediates or an early step during Vipp1 ring degradation. In either case the strings clearly indicate an organization of Vipp1 by lateral accumulation of multiple monomers. Occasionally, two Vipp1 rings can further interact to form a double-ring structure (Figure 9, D–H). It is appealing to speculate that both Vipp1 rings interact with a rather hydrophobic surface, which is otherwise involved in membrane binding. Single-particle EM analysis of purified Vipp1 showed formation of various homooligomeric Vipp1 ring structures with different diameters. Actually, Vipp1 is to our knowledge rather special to make such a variety of rings, which means that there must be a special function linked to it, as will be discussed below. A similar diversity of ring structure with different compositions of the building units has - to our knowledge - only been described so far for the IsiA complex from *Synechocystis* (Boekema *et al*., 2001; Kouril *et al*., 2003; Kouril *et al*., 2005; Yeremenko *et al*., 2004)

It is also interesting to compare the Vipp1 rings with those of the structural and functional homologous PspA protein. The Vipp1 rings have a 12–17-fold symmetry and a diameter of 25–33 nm, whereas the PspA rings are uniform with a ninefold symmetrical

configuration (Hankamer *et al*., 2004). If we extrapolate to a hypothetical Vipp1 ring with nine fold symmetry, this ring would have a diameter of 22 nm. This is close to the actual PspA ring diameter of 20 nm, if we take into account that the latter ring has a rather smooth surface because it lacks spikes like Vipp1 has. The PspA rings were considered to be composed of tetrameric building blocks to form uniform rings with a molecular mass of ~1 MDa (Hankamer *et al*., 2004). This equals the size of half of the Vipp1 rings, which can be explained by the fact that the ring height is only 8.5 nm. This difference in the height of the PspA rings versus the Vipp1 rings indicates that packing of the PspA and Vipp1 monomers in the two rings must be different. Although Vipp1 is C-terminally prolonged when compared with PspA, oligomerization of the Vipp1 monomers depends only on the PspA-like N-terminal domains (Aseeva *et al*., 2004). Thus, the C-terminal extension of Vipp1 is most likely not directly involved in formation of the ring structures. Because of this, the N-terminal core domain of PspA and Vipp1 must pack differently. Although only ~25% of the PspA amino acid residues must form an extended α-helix to span the height of the PspA ring (8 nm), ~50% of Vipp1 has to form an extended α-helix to span 22 nm. For both proteins a domain of ~140 amino acids of the N-terminal region is predicted to form an α-helix with a high propensity for coiled-coil formation. For Vipp1 this predicted α-helical domain of 140 amino acids would have the right length to account for the determined height of the rings of 22 nm, if present as one giant extended α-helix. In the case of PspA this helix is most likely folded as a hairpin and may form an α-helix bundle that would result in an increase of the PspA ring width. Different structural organization of PspA and Vipp1 might also indicate small differences in the respective protein functions, which would, e.g., explain why *Synechocystis* contains PspA as well as Vipp1.

It has to be noted that we did not observe cage-like structures as described before for the PspA from *E. coli*. Because for the preparation of the *Synechocystis* Vipp1 no detergent was used the suggested major impact of detergent on the PspA structure (ring vs. cage structures) should be revisited.

The observed diverse ring structures of Vipp1 could be involved in vesicle formation in cyanobacteria and chloroplasts. It is intriguing to speculate that Vipp1 assembled into a ring around the neck of a forming budding vesicle, similar to dynamin in eukaryotic cells.

Stepwise release of individual Vipp1 units results in formation of rings with smaller diameters, which would bring the membranes into a close contact, resulting in membrane fusion and pitching off of a vesicle. If Vipp1 performs such a dynamin-like function, the protein will most likely recruit other proteins that are, for example, involved in initial membrane bending and vesicle fission. Proteins involved in regulating the controlled

Vipp1 (dis)assembly could involve Hsp70 chaperones and co-chaperones, as discussed recently (Liu *et al*., 2007).

Taken together, we show here that *Synechocystis* Vipp1 forms various oligomeric rings with different monomer stoichiometries. Although some of these rings exist in the cytoplasm of *Synechocystis* in a soluble form, most Vipp1 oligomers are tightly attached to both the thylakoid and the cytoplasmic membrane of the cyanobacterium *Synechocystis* PCC 6803. Formation of promiscuous Vipp1 ring structures could play a critical role in vesicle formation in cyanobacteria and chloroplasts.

Acknowledgement

We thank Claudia Escher for technical assistance and Mathias Weber for help with the GALLEX measurements. This work has been supported by a grant from the Deutsche Forschungsgemeinschaft (SCHN 690/3-1) and the Council of Chemical Sciences (C.W.) of the Netherlands Science Foundation of the Netherlands Organization for Scientific Research.

Chapter 4

Vipp1 and PspA: related but not twins

Jelle B. Bultema, Eva Fuhrmann, Egbert J. Boekema and Dirk Schneider

This chapter has been partly published in *Communicative and Integrative Biology* (2010) 3

Abstract

The *Vesicle Inducing Protein in Plastids 1* (Vipp1) has been suggested to be involved in thylakoid membrane formation in both chloroplasts and cyanobacteria. The protein shows sequence homology to the *Phage Shock Protein A* (PspA) from bacteria, and both proteins have similar secondary structures. 2D-structures of PspA and of Vipp1 have been determined by electron microscopy in recent years. Both PspA and Vipp1 form large homooligomeric rings with high molecular masses but their ring dimensions differ significantly. Furthermore, Vipp1 forms rings with different rotational symmetries whereas PspA appears to form rings with a single rotational symmetry. In this chapter we compare the structures of PspA and Vipp1 and suggest a spatial structural model of the observed Vipp1 rings.

Introduction

Thylakoid membranes in green plants, algae and cyanobacteria house the proteins which are capable to carry out oxygenic photosynthesis. In cyanobacteria, the thylakoid membranes are organized as ring-like layers in an onion and often lay parallel to the plasma membrane following the shape of the cell. However, there is a known exception to this rule: the species *Gloeobacter violaceus*. This primitive cyanobacterium lacks any form of internal thylakoid membrane (Koenig *et al*., 1995) and the photosynthetic apparatus is located in the plasma membrane. In plant cells the thylakoid membranes are lipid bilayers which are inside the chloroplast organelle. The thylakoid membranes can either be stacked which are called grana or unstacked which are known as stroma lamellae. One can speculate about the origin of thylakoid membranes in both cyanobacteria and plants. Possibly at one point during evolution the plasma membrane from a cyanobacterium like *Gloeobacter violaceus* started to protrude into the cytoplasm. The resulting membrane foldings became packed with photosynthetic proteins and separated from the plasma membrane. By such a type of differentiation, the newly formed membranes became thylakoid membranes as found in most cyanobacteria (Vothknecht *et al*., 2001). Chloroplasts are nowadays seen as having originated from a prokaryote with photosynthetic activity invading a protoeukaryotic host cell whereby the prokaryote evolves into its current form as an organelle (Schwartz *et al*., 1978). This event is known as endosymbiosis.

The mechanism behind the formation of thylakoid membranes in both cyanobacteria and plants is nowaday not yet fully understood. It was shown that light can induce thylakoid membrane formation (Tzinas *et al*., 1987), but the underlying biochemical processes are elusive. However, it was found in 2001 that deletion of gen *hcf155* from *Arabidopsos thaliana* resulted in the total loss of thylakoid membranes. This implied a possible role for the product of *hcf155* during chloroplast development.(Kroll *et al*., 2001). Moreover, the same *Arabidopsis* mutations lost the possibility to form vesicles, hence the name Vesicle Inducing Protein in Plastids 1 or Vipp1.

Homologues of the chloroplast Vipp1 were found in the genomes of several cyanobacteria. While *vipp1* could not be completely deleted in the cyanobacterium *Synechocystis* PCC 6803 (no entire segregation could be achieved, suggesting that loss of Vipp1 is lethal to the cells), depletion of the gene product resulted in a severe reduction of internal thylakoid membranes (Westphal *et al*., 2001). Thus the protein appears to have similar function in cyanobacteria and chloroplasts. Another observation supporting this was that depletion of Vipp1 also resulted in a decreased level of photosystems in both chloroplasts (Aseeva *et al*., 2007) and cyanobacteria (Fuhrmann *et al*., 2009b; Gao *et al*., 2009). So, these observations allow us to say that it is very likely that Vipp1 is at some point involved in the thylakoid membrane biogenesis, although the exact function of Vipp1 is still not known.

The sequence of Vipp1 shows a high degree of sequence similarity to the phage shock protein A (PspA) from bacteria. Expression of PspA was initially identified to be highly upregulated after phage infection of *E. coli* cells, but it was shown later that expression of PspA appears to be generally increased under stress conditions (Brissette *et al*., 1990). The gene encoding PspA can be deleted in *E. coli* without a dramatic phenotype. It was also shown that the protein binds to specific lipids of the membrane (Kobayashi *et al*., 2007). Furthermore, blocking the fatty acid biogenesis in *E. coli* resulted in cells with a detached cytoplasmic membrane, but more interesting, also in an increased production of PspA. This is supported by more recent observations that indicate that the protein is involved in maintaining the integrity of the *E. coli* cytoplasmic membrane (Standar *et al*., 2008). Vipp1 associates tightly with both the cytoplasmic or inner envelope and the thylakoid membrane in cyanobacteria and chloroplasts (Fuhrmann *et al*., 2009b; Kroll *et al*., 2001; Li *et al*., 1994).

Interestingly, many cyanobacteria contain both a *pspA* as well as a *vipp1* gene. As in *E. coli*, the single *pspA* gene of the cyanobacterium *Synechocystis* can be deleted without resulting in a dramatically changed phenotype (Weiner *et al*., 1994) whereas depletion of the *vipp1* gene product resulted already in severe defects (Aseeva *et al*., 2007; Fuhrmann *et*

al., 2009b; Kroll *et al*., 2001; Westphal *et al*., 2001). These observations indicate that Vipp1 is essential in cyanobacteria as well as in chloroplasts whereas PspA is not crucial in bacteria. Expression of the *Synechocystis* Vipp1 in *E. coli* could complement the deletion of the endogenous *E. coli* pspA gene under certain stress conditions. This observation indicates that the two proteins could have similar functions in vivo. Interestingly, the observation that many cyanobacteria encode both PspA and Vipp1 raises the question of how the functions of these two related proteins are separated in these organisms. Noteworthy, in some cyanobacteria, such as *Microcystis aeruginosa* and *Acaryochloris marina*, only a PspA homologous protein is encoded and no Vipp1. This indicates that in these strains either PspA also fulfils the function of Vipp1 or that in these strains thylakoid membrane biogenesis is organized differently without involvement of a Vipp1 homologous protein. Also the cyanobacterium *Gloeobacter violaceus* lacking thylakoid membranes (Koenig *et al*., 1995), does not encode Vipp1 but only its homologue PspA.

In 2004 the 2D structure of PspA from *E. coli* was determined by electron microscopy (EM) (Hankamer *et al*., 2004). The protein forms a large homooligomeric ring of about 1 MDa. The ring has a 9-fold symmetry and it was calculated that the ring contains 36 copies of the PspA monomer per ring. Very recently, we published a structural analysis of Vipp1 from *Synechocystis* (Fuhrmann *et al*., 2009a), as discussed in the previous chapter of this thesis. In line with the PspA structure, also Vipp1 forms large homooligomeric ring structures. However, Vipp1 was found to form diverse higher ordered ring structures with 12-17 fold symmetry, in contrast to the pure population of 9-fold symmetrical PspA rings. The higher symmetry of Vipp1 also leads to a higher mass of the protein which is about 1.8-2.2 MDa. The differences in the structures of PspA and Vipp1 could correlate with their defined physiological roles.

Here, we used bioinformatics and modelling for a detailed comparison between Vipp1 and PspA from *Synechocystis* and PspA from *E. coli* at all structural levels. Starting with the primary structured we end with a predicted 3D model for the Vipp1 multimeric organization.

Structural comparison of Vipp1 and PspA: Primary structure

The amino acid sequence of Vipp1 and PspA are known to be highly homologous. The protein sequences of PspA from *E. coli* as well as of PspA and Vipp1 from *Synechocystis* were aligned as shown in figure 1. All three proteins align from the N-terminus up to amino acid (aa) residue 218 without insertions or deletions. The two PspA proteins share a sequence identity of only 27% but show a sequence similarity of 51%. These data

indicate already that the PspA proteins and homologues are not highly conserved. The *Synechocystis* PspA protein and *Synechocystis* Vipp1 also share just 31% sequence identity (51% similarity). Taken together, the amino acid sequence conservation between PspA and Vipp1 homologs is low. Furthermore, Vipp1 contains a C-terminal extension of 45 amino acids, and this extension appears to be the characteristic feature of Vipp1 vs. PspA proteins.

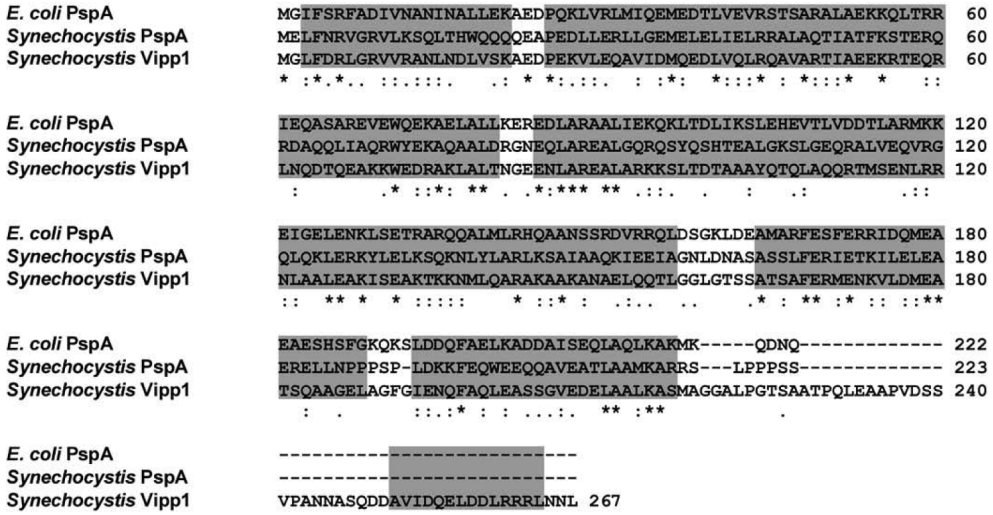


Figure 1. Amino acid sequence alignment of Vipp1 from *Synechocystis* and PspA from *Synechocystis* and *E. coli* with Clustal X.14 (Thompson *et al*., 1997). Identical (*), strongly similar (:) and weakly identical (.) amino acid residues are indicated. The grey regions represent α-helical domains predicted with PSIPRED, V2.6.15 (Larkin *et al*., 2007).

Structural comparison of Vipp1 and PspA: Secondary structure

Although PspA and Vipp1 show some degree of sequence identity, the amino acid sequence of the proteins is not highly conserved. However, the proteins have a more conserved secondary structure. The predicted secondary structure for all proteins is essentially purely alpha-helical, apart from some coil interruptions. This prediction is in excellent agreement with experimental determination of the PspA and Vipp1 secondary structure (Hankamer *et al*., 2004). The grey highlighted amino acids in figure 1 represent the regions predicted to be α-helical. In addition to the regions within the PspA domain of Vipp1, this protein contains an extra α-helical domain at its C-terminus. This additional domain appears to be the unique feature of Vipp1 when compared to PspA and it is not involved in oligomer formation, as has been shown recently (Aseeva *et al*., 2004).

The program PCOILS (Lupas, 1996) was used to determine the probability of Vipp1 and PspA to form coiled-coils of helices. Analysis revealed that the first half of both proteins can form coiled-coils, for Vipp1 these regions are amino acids 25-60 and 80-160 and for PspA amino acids 30-150 and 160-185. Thus, the predicted secondary structures of pspA and Vipp1 are conserved as well as the probability of the protein domains to form coiled-coils.

Structural comparison of Vipp1 and PspA: ring vs. rings

It was shown in several electron microscopy studies that Vipp1 and PspA can form large ring structures (Vipp1: Aseeva *et al*., 2004; Fuhrmann *et al*., 2009a), (PspA: Hankamer *et al*., 2004; Standar *et al*., 2008). The structural details of these rings, including dimensions and observed symmetry are shown in table 1. From this table it is striking that the analogy in primary and secondary structures does not hold for the quaternary structure where a rather large variation for ring-dimensions exists. Furthermore, the studies on PspA are in disagreement with each other concerning the dimensions and the same is true for the two studies on Vipp1. The discrepancy is that half of the studies mention a ring population with distinct dimensions (Aseeva *et al*., 2004; Hankamer *et al*., 2004) whereas the other half describes variable dimensions for these ring structures (Fuhrmann *et al*., 2009a; Standar *et al*., 2008). This remarkable observation can be explained by the different purification procedures. Usage of detergent at some point during the purification procedure results in a ring population with distinct dimensions whereas detergent-free protein preparations results in variable rings dimensions. This suggests that Vipp1 and PspA are influenced by the detergent in such a way that only one particular specific quaternary structural arrangement is present.

Table 1. A comparison of structure and symmetry of rings of PspA and Vipp1, as determined by electron microscopy and image analysis.

	Width (nm)	Height (nm)	Symmetry	Copies/ring	Detergent	Reference
PspA	20	8.5	9 fold	36	Yes	(Hankamer *et al*., 2004)
	24.9-56.2	-	-	-	No	(Standar *et al*., 2008)
Vipp1	40	14	-	-	Yes	(Aseeva *et al*., 2004)
	25-33	22	12-17 fold	48 - 68	No	(Fuhrmann *et al*., 2009a)

Tertiary structure prediction and quaternary structure modelling for Vipp1

So far, no crystal structures are available for Vipp1 or for PspA. However, an impression of the hypothetical tertiary structure can be obtained by 3D-structure prediction. For this purpose the sequences of the *Synechocystis* Vipp1 as well as of the *E. coli* PspA were submitted to the Phyre webserver (Kelley *et al*., 2009). All predicted 3D models for Vipp1 and PspA had either one long alpha-helix or coiled-coil structures. According to Kelly and Sternberg, when the predicted models have the highest estimated precision but rather low sequence identity between input protein and the proteins used as starting models, this means that the core of the models will tend to be accurate but peripheral regions could show greater deviations. Taking this into account, the model based on the 3D structure of smooth muscle alpha-actinin (pdb entry 1sjj) was chosen for modelling of the oligomeric ring structure of Vipp1. This model (figure 2) contains amino acids 4-253 from Vipp1 which equals 93% of the whole Vipp1 sequence.

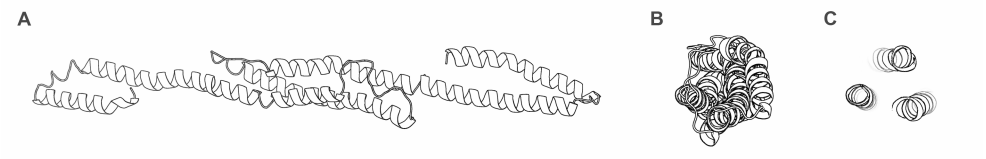


Figure 2. Modelling of the monomeric Vipp1 structure. (A) Ribbon presentations from aside and (B) from the bottom, which is equivalent to the left side in the left frame. (C) presents a cross-section at the central part of the model. Pymol was used for visualisation of all models.

The monomeric Vipp1-model has a length of 17 nm and a diameter of ~2.5 nm. Partly, it is a three-start coiled-coil in the central part of the structure as shown in figure 2C. Already at this stage it becomes obvious that the tertiary structure of PspA and Vipp1 must differ significantly. The suggested Vipp1 model has a length of 17 nm which corresponds to a ring height of >17 nm (for comparison, see below). Since the PspA ring has only a height of 8.5 nm (Hankamer *et al*., 2004), the generated homology model cannot be used to derive the 3D structure of the PspA ring, and already the tertiary structure of PspA must significantly differ from Vipp1. It is e.g. possible that in the PspA structure the long α-helices are strongly kinked to form a higher-ordered helix-bundle with a length of about 8-9 nm.

Based on biochemical characterisation it was suggested that the smallest building-block within the Vipp1 ring structures is a dimer of dimers. Single particle analysis of *Synechocystis* Vipp1 top-views (Fuhrmann *et al*., 2009a) showed that the ring structures have a rotational symmetry ranging from 12- up to 17-fold. Ring structures with 15-fold

symmetry are in the middle of the observed range of symmetries. This type of ring has 15 spikes pointing outwards and 15 pointing to the inside. To each spike two densities can be attributed, which both have a diameter of ~2.5 nm. The fact that the diameters of the observed 2.5 nm densities and of the predicted Vipp1 model are about the same, leads us to speculate that the density could originate from a single Vipp1 protein copy. From this we conclude that the spikes, as seen in the EM projection maps, are each occupied by two Vipp1 proteins. Two predicted Vipp1 models were arranged manually such that the total dimer length was 22 nm, which corresponds to the ring height determined by single particle analysis (Fuhrmann *et al*., 2009a). Furthermore, two dimers were placed next to each other based on the spikes in the EM projection map with 15-fold symmetry (figure 3B). The 15-fold ring model (figure D-F below) was generated from this tetramer using routines from the CCP4 package (Collaborative Computational Project Number 4, 1994).

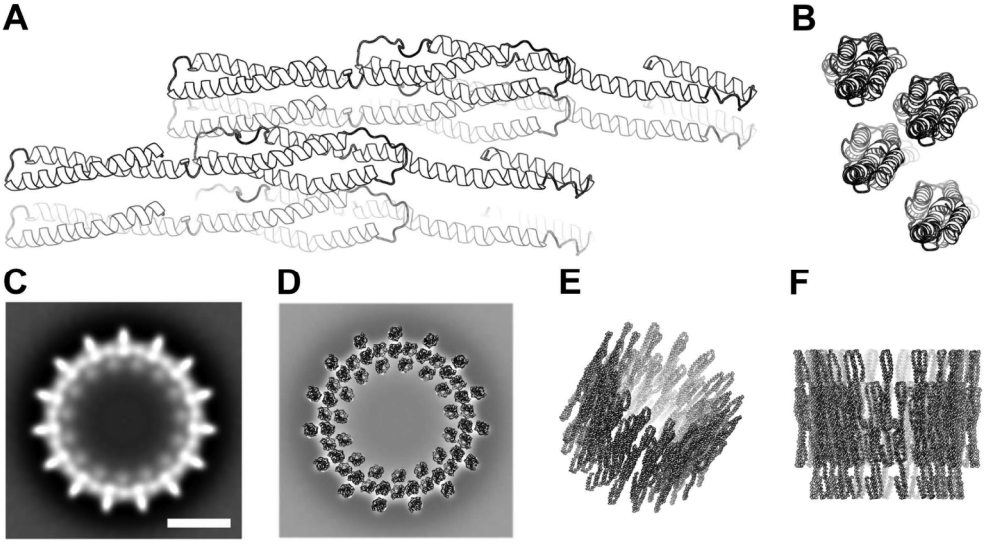


Figure 3. Model of the Vipp1 structure. (A, B) Side and top view ribbon representations of four predicted Vipp1 monomers arranged in a dimer of dimers configuration. (C) Projection map showing a Vipp1 ring with 15-fold symmetry. (D-F) Top, tilted and side view of the modelled Vipp1 ring, respectively. In (D) the model is placed on top of the projection map shown in (C). The scale bar in (C) equals 10 nm which is also valid for (D-F).

The presented ring model provides a good impression of how the Vipp1 ring structure could look like. Most importantly, the α-helical configurations of the model are in good agreement with the overall ring dimensions.

Conclusions/Future directions

In summary, Vipp1 and PspA have a highly similar secondary structure despite a rather low conservation of the primary sequence. Although no high-resolution tertiary or 3D quaternary structure is available for Vipp1, we could generate a 3D ring model which is consistent with the about 2 nm resolution electron microscopy projection maps as well as with the structure predictions. This model is in good agreement with dimensions of Vipp1 (Fuhrmann *et al*., 2009a) but is not in line with the solved PspA structure (Hankamer *et al*., 2004). Keeping in mind the dimensions of the 3D structure of its complex, the 3D structure of monomeric PspA must be more compact. Its predicted alpha-helices and coiled coils cannot be that extended as suggested here for Vipp1. The difference between the tertiary structures of PspA and Vipp1 could be a explanation why both proteins can coexist in certain cyanobacteria without interfering with each other. Most likely, the two proteins can only form homo-oligomers, although with a remarkably similar coarse structure.

In some cyanobacteria only a PspA homologous protein exists but no Vipp1 homolog. It would be intriguing to determine the structure to see if the protein has a PspA or a Vipp1 structure.

While both PspA and Vipp1 form large homo-oligomeric ring structures, the physiological importance of the oligomerization event is unclear. Do the rings represent an assembly intermediate or only a storage form, or is the ring essential for the PspA/Vipp1 function? A C- and N-terminal truncation study with the myotonic dystrophy protein kinase (DMPK) showed that its coiled-coil domain modulates oligomerization and that the oligomer is involved in substrate binding (Herpen *et al*., 2006). Since Vipp1 and PspA both have two large predicted coiled-coil domains, this could be an explanation for the oligomerization into ring structures.

Last but not least the question arises if the central hole found in the PspA and Vipp1 structure is of structural importance. Is it important for Vipp1 to form different rings with different diameters of the central hole?

Acknowledgements

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SCHN 690/3-1) and the Council of Chemical Sciences (CW) of the Netherlands Science Foundation NWO. We thank Marcel Bokhove (Department of Biophysical Chemistry, University of Groningen) for help with modelling.

Chapter 5

Structural basis for CRISPR RNA-guided recognition of DNA by Cascade

Matthijs M. Jore*, Magnus Lundgren*, Esther van Duijn*, Jelle B. Bultema*, Edze R. Westra, Sakharam P. Waghmare, Blake Wiedenheft, Marieke R. Beijer, Arjan Barendregt, Kaihong Zhou, Ambrosius P.L. Snijders, Mark J. Dickman, Jennifer A. Doudna, Egbert J. Boekema, Albert J. R. Heck, John van der Oost & Stan J.J. Brouns

* These authors contributed equally to this work

Manuscript in preparation

Abstract

The CRISPR immune system in prokaryotes utilizes small guide RNAs to neutralize invading viruses and plasmids. In *Escherichia coli*, immunity is dependent on a ribonucleoprotein complex called Cascade. Here we present the composition and low-resolution structure of Cascade and show how it recognizes double-stranded DNA targets. Cascade is a 405 kDa complex comprising five Cas proteins (CasABCDE, 1:2:6:1:1) and a 61 nucleotide crRNA. The crRNA guides Cascade to the complementary sequence of double-stranded invader DNA by ATP-independent strand displacement. Although DNA recognition does not require a protospacer adjacent motif in vitro, the motif is essential for virus resistance in vivo, as are all Cascade components. CasA enhances target DNA recognition while CasCDE constitutes the core of the complex. The structure of Cascade reveals an unusual seahorse-shape that undergoes conformational changes upon target DNA binding. A structural model is presented providing insight into the molecular basis of crRNA-guided target DNA recognition.

Introduction

The constant pressure of invading viruses and conjugative plasmids has shaped the evolution of host defense systems in prokaryotes. The widely distributed CRISPR (clustered regularly interspaced short palindromic repeats) immune system represents the most recently discovered prokaryotic defense strategy (reviewed by (Horvath *et al*., 2010; Karginov *et al*., 2010; van der Oost *et al*., 2009)). The system consists of repeats that are interspaced by unique sequences called spacers, which are derived from viral and plasmid DNA (Bolotin *et al*., 2005; Mojica *et al*., 2005; Pourcel *et al*., 2005). CRISPR-based immunity is adaptive and inheritable because it can both memorize invaders by storing fragments of their DNA, and pass that information on to subsequent generations following a Lamarckian type of evolution (Koonin *et al*., 2009; van der Oost *et al*., 2009).

The CRISPR-associated (Cas) protein machinery is encoded by gene clusters that are located in close proximity of the CRISPR locus (Jansen *et al*., 2002), which has allowed for the extensive horizontal transfer of complete CRISPR/Cas systems (Godde *et al*., 2006). Multiple types of *cas*-gene sets have been recognized (Haft et al., 2005; Makarova et al., 2006) that correlate with specific families of repeat sequences (Kunin et al., 2007).

The mechanism of CRISPR/Cas-induced immunity has been divided into three stages. In the first stage, CRISPR adaptation, the host encounters an invader and integrates a new spacer derived from the foreign DNA into the CRISPR, resulting in resistance to foreign genetic elements carrying this sequence (Barrangou *et al*., 2007; Horvath *et al*., 2008; van der Ploeg, 2009). Although the molecular determinants of foreign DNA recognition have not been elucidated, the crystal structure and metal-dependent DNase activity of Cas1 indicated that this enzyme maybe involved in generating small DNA fragments that are used as precursors for CRISPR adaptation (Wiedenheft *et al*., 2009). Newly acquired spacers from both the coding and template strand of the viral genome have been shown to confer immunity (Barrangou *et al*., 2007)

In the second stage, CRISPR expression, the CRISPR locus is transcribed and the repeat regions within the precursor CRISPR RNA (pre-crRNA) are cleaved (Hale *et al*., 2008; Lillestøl *et al*., 2009; Tang *et al*., 2002) by a specific Cas endoribonuclease. Two pre-crRNA processing endonucleases have been described: CasE from *Escherichia coli* (Brouns *et al*., 2008) and Cas6 from *Pyrococcus furiosus* (Carte *et al*., 2008). In *E. coli*, CasE is an essential component of the Cascade complex (CRISPR-associated complex for antiviral defense), which consists of five Cas proteins from the Cse-type (CasABCDE) (Fig.1a). CasE-generated mature crRNAs remain bound to Cascade to guide host defense. This is distinct from the situation in *P. furiosus* in which Cas6-generated crRNAs end up in the Cmr-complex encoded by the RAMP module of Cas proteins, where they are further trimmed at the 3'-end from ~67 to 39 or 45 nucleotides. The guide RNA-loaded Cmr-complex cleaves single stranded target RNA sequence-specifically (Hale *et al*., 2009). In *E. coli*, the third stage, CRISPR interference, not only requires Cascade loaded with anti-invader crRNA, but also the participation of the predicted nuclease/helicase Cas3. Because crRNAs complementary to the either strand of the phage DNA provided resistance, it was proposed that Cascade is a crRNA guided complex that targets DNA rather than mRNA (Brouns *et al*., 2008). A series of genetic experiments in *Staphylococcus epidermidis* involving a conjugative plasmid showed that the Csm-type CRISPR/Cas system targets DNA as well (Marraffini *et al*., 2008). Although DNA is the prime candidate for the target molecule in the *E. coli* and *S. epidermidis* model systems, direct molecular evidence of Cas proteins interacting with their target DNA has been lacking. In this study we show how Cascade recognizes target DNA and present a structural model of Cascade that provides insight into the mechanism of crRNA-guided recognition of DNA targets.

Results

Core sub-complexes of Cascade

The *E. coli* K12 CRISPR/Cas system (Cse-subtype) consists of a gene cluster encoding eight *cas* genes (*cas3*, *casABCDE*, *cas1* and *cas2*) and a downstream CRISPR locus (Fig. 1A). Five Cas proteins (CasABCDE) form the Cascade complex, which cleaves a long precursor transcript of the CRISPR region (pre-crRNA) into small crRNA molecules. These crRNAs remain bound to the complex to guide antiviral defense (Brouns *et al*., 2008). To investigate the role of the individual subunits, we first tested whether each subunit is required for antiviral defense. Viral plaque assays with Cas3 and Cascade lacking one type of protein subunit showed that all protein components of Cascade are indispensible for the virus resistant phenotype of *E. coli* (not shown).

We then systematically overproduced and affinity purified different combinations of Cascade subunits and checked for the presence of mature crRNA. This analysis showed that CasA or CasAB could be omitted without affecting the apparent stoichiometry of the remaining subunits or the mature crRNA (Figs. 1B and 1C). We noticed that Cascade, unlike CasBCDE and CasCDE, always co-purified with large nucleic acid molecules (>300 nt) (Fig. 1C). Removal of the Cas proteins followed by nuclease treatments showed that RNase A only hydrolyzed the crRNA, while DNase I removed the long nucleic acids, thereby identifying the co-purified nucleic acid as DNA (Fig. 1D). Size exclusion chromatography of the three types of complexes revealed that the vast majority of CasBCDE and CasCDE were present in a single form, whereas Cascade displayed a substantial void peak in addition to a discrete peak at ~11 ml (Fig. 1E). DNase I treatment prior to gel filtration eliminates the void peak without disruption of the discrete Cascade peak, again indicating the presence of Cascade-bound DNA (Fig. 1E).

Architecture of crRNA

The characteristics of the mature crRNA species were accurately determined by subjecting mature crRNAs isolated from Cascade to denaturing RNA chromatography (Dickman *et al*., 2006; Waghmare *et al*., 2009) and electrospray ionization mass spectrometry (ESIMS). To simplify the analysis, a uniform crRNA preparation was obtained by co-expressing Cascade with a designed CRISPR containing eight repeats and seven identical spacers (denoted R44 CRISPR (not shown)). This setup resulted in a Cascade preparation in which each molecule was loaded with the same crRNA. Chromatography demonstrated the purity and homogeneity of this crRNA preparation (Fig. 1F).

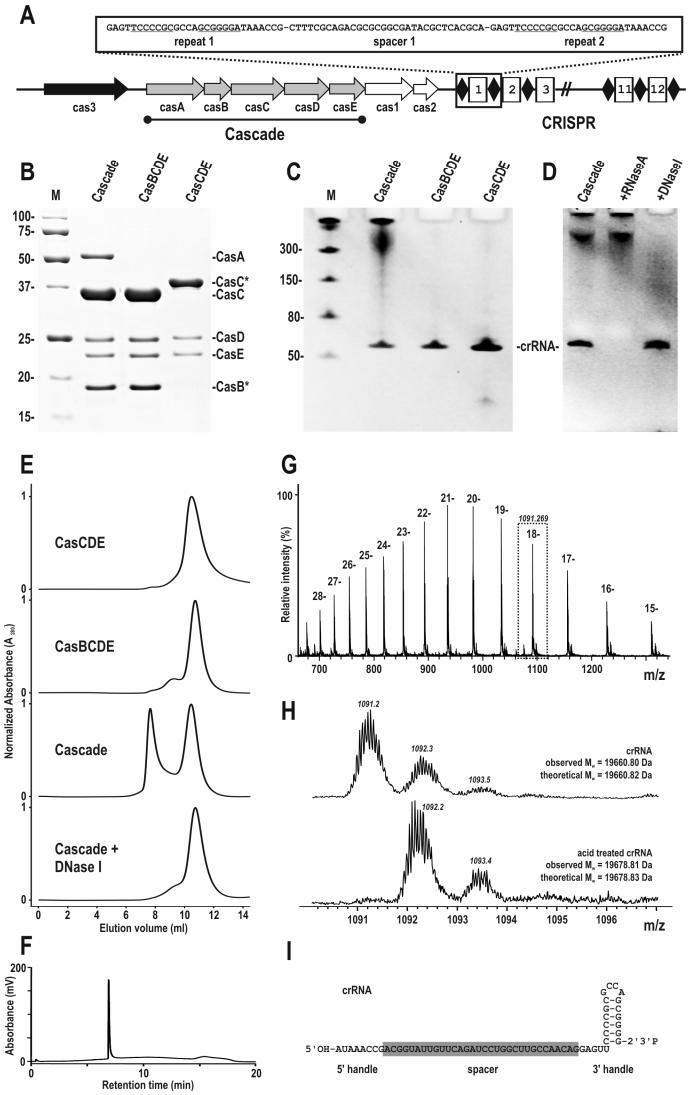


Figure 1. Core complexes of Cascade retain crRNA of 61 nt. A) Schematic diagram of the CRISPR/Cas locus in *E. coli* K12 containing *cas3* (*ygcB*), *casA* (*cse1*, *ygcL*), *casB* (*cse2*, *ygcK*), *casC* (*cse4*, *ygcJ*), *casD* (*cas5e*, *ygcI*), *casE* (*cse3*, *ygcH*), *cas1* (*ygbT*), *cas2* (*ygbF*) [Haft et al., 2005; Jansen et al., 2002]. B) Coomassie blue-stained SDS-polyacrylamide gel showing StrepTactin purified Cascade, CasBCDE and CasCDE. Protein marker sizes in kDa. The asterisk marks the Strep-tagged subunit. C) Ethidium bromide-stained denaturing PAA-gel showing nucleic acids isolated from purified Cascade (sub-)complexes. RNA marker sizes in nucleotides. D) RNase A or DNase I treatment of Cascade bound nucleic acids from Cascade. E) Size exclusion elution profiles of CasCDE, CasBCDE and Cascade before and after DNase I treatment. F) Ion-pair reversed-phase HPLC purification of mature R44 crRNA (not shown) at 75 °C. G) Multiple charged ESI-MS spectra of the purified mature crRNA. H) Enhanced view of the 18- charged species before (upper graph) and after acid treatment (lower graph) indicating the hydrolysis of the 2',3'-cyclic phosphate. I) diagram of mature crRNA derived from the R44 CRISPR showing the 5' hydroxyl group and 2',3'-cyclic phosphate.

Furthermore, the observed retention time was consistent with an approximate length of 60 nt. The ESI-MS spectra indicated that the crRNA had a molecular weight of 19,660.80 Da (Fig. 1G), which corresponds well to an expected molecular weight of 19,660.82 Da for a 61 nt crRNA resulting from a single CasE endoribonuclease cleavage event in each repeat. The purified mature crRNA was also analyzed using ESI-MS/MS analysis following RNase T1 and RNase A digestion. A number of oligoribonucleotide digests were assigned to the mature crRNA sequence (not shown) and were consistent with the previously determined CasE cleavage site, 5' of the terminal base of the hairpin (Brouns *et al*., 2008). The molecular weight analysis of the crRNA indicated a 5'- hydroxyl group and a 2',3'-cyclic phosphate terminus. The presence of a cyclic phosphate terminus was confirmed by acid treatment of the crRNA, which showed a mass shift of 18 Da, corresponding to the hydrolysis of the 2',3'-cyclic phosphate to a 2' or 3' phosphate (Fig. 1H). Mature crRNA is 61 nucleotides long and contains the 32 nucleotide spacer sequence, flanked by repeat-derived sequences on either end: 8 bases at the 5' terminus (5'-handle) and 21 bases forming a hairpin with a tetra-nucleotide loop at the 3' terminus (3'-handle) (Fig. 1I).

Target recognition by Cascade

The observation that DNA co-purified with Cascade (Figs. 1C and 1D) prompted us to analyze the DNA binding behavior of Cascade in detail. Electrophoretic mobility shift assays (EMSAs) demonstrated that Cascade was able to bind single-stranded (ss) DNA containing the protospacer, a sequence complementary to the spacer sequence of the crRNA (Figs. 2A, 2C and S4). Double-stranded (ds) target DNA was also bound sequence-specifically, without the need for additional co-factors such as divalent metal-ions or ATP (Figs. 2B and 2D). The dissociation constant (Kd) of Cascade for single- and double-stranded target DNA was 8 and 790 nM, respectively. In addition to target DNA, Cascade also bound non-target DNA, i.e. DNA without a protospacer, but this interaction was much weaker (Figs. 2C and 2D). Competitor DNA blocked Cascade from binding non-target DNA, and at very high competitor concentrations also from binding target DNA (Figs. 2A-D). The competitor had little effect on preformed Cascade-target DNA complexes, indicating rapid irreversible interaction between Cascade and complementary DNA substrates (not shown).

Cascade sub-complexes lacking CasA (CasBCDE and CasCDE) displayed only sequence-specific binding to ss- and dsDNA targets, and did not bind non-target DNA

(Figs. 2E-H). Consistent with that observation, the target DNA binding behaviour of these subcomplexes was not affected by the addition of competitor DNA.

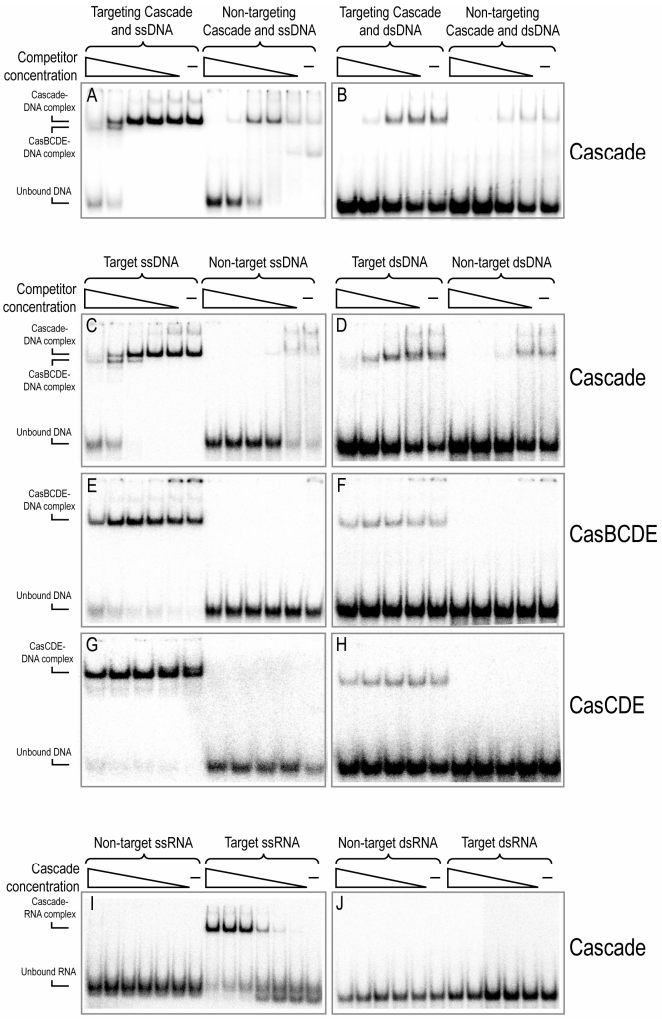


Figure 2. Target recognition by Cascade. A-B) Effect of the type of crRNA bound. Cascade was loaded with either targeting crRNA (derived from the R44 CRISPR) or non-targeting crRNA (derived from the K12 CRISPR). The binding of these two types of Cascade complexes to one type of probe is shown. DNA probes are 85 nucleotide ssDNA or dsDNA sequences containing the R44 protospacer (32 nucleotides), flanked by 27 nucleotides on either end CH) Effect of uniform crRNA-loaded complexes (R44 CRISPR) on the binding of DNA probes with and without protospacer (denoted target and non-target, respectively). The binding of Cascade, CasBCDE and CasCDE to an 85 nucleotide target and non-target ssDNA and dsDNA is shown. Non-target DNA probes contain a sequence-scrambled R44 protospacer. I-J) Effect of uniform crRNA-loaded Cascade (R44 CRISPR) on the binding of target and nontarget ssRNA and dsRNA. A-H) DNA competitor concentration is 2500, 500, 50, 5 and 0.5 ng/µl (*n.b.* the highest concentration was not used for CasCDE), and protein concentration 70- 125 ng/µl, and in I-J the Cascade concentration range is 160, 80, 40, 20, 10 and 5 ng/µl.

The addition of purified CasA to CasBCDE preparations restored Cascade-like non-specific DNA binding ability (not shown), but neither CasA alone (not shown) nor the combination CasA and CasCDE (not shown) displayed non-specific DNA binding. Because the mobility shift caused by Cascade or CasBCDE binding could be distinguished (Figs. 2A, 2C and S4), competition assays were performed between the two types of complexes. This showed that the simultaneous addition of equal amounts of both Cascade and CasBCDE resulted in substantially more target DNA binding by Cascade. An approximately 25-fold higher CasBCDE concentration was required for the protein complexes to shift equal amounts of target (Fig. 3A). Furthermore, less dsDNA 9 target was shifted by CasBCDE than Cascade at equivalent conditions (Figs. 2D and 2F), again indicating that CasA enhances target DNA localization.

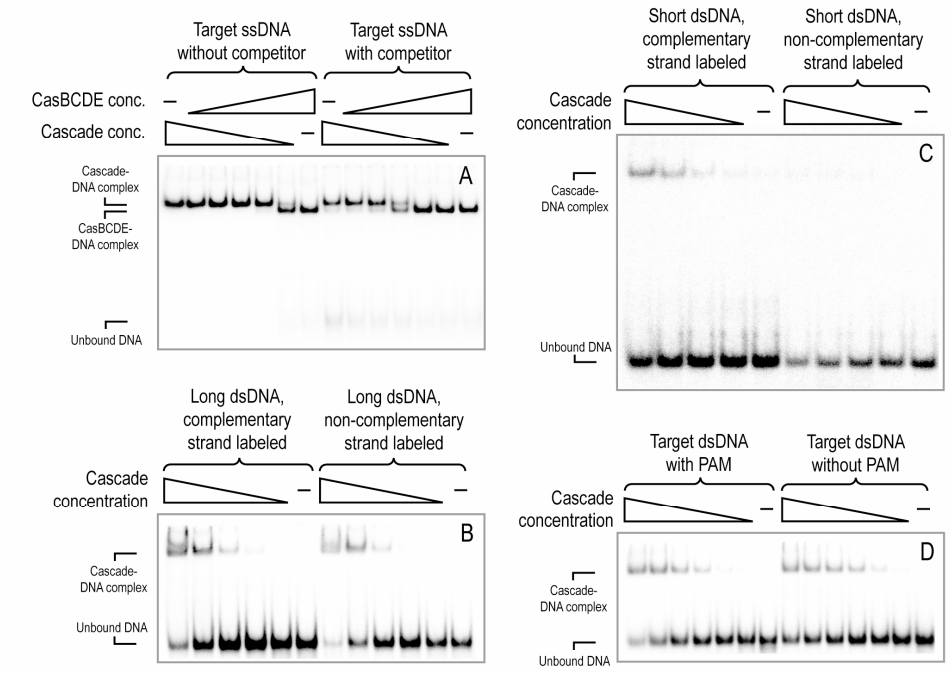


Figure 3. A) Competition assay between R44 crRNA-loaded Cascade and CasBCDE for R44 ssDNA target. A total of 160-180 ng of protein used in each reaction in a Cascade:CasBCDE ratio of 1:0, 99:1, 9:1, 1:1, 1:9, 1:99 and 0:1, respectively. B) Effect of labeling the complementary or non-complementary strand of a long dsDNA target containing the R44 protospacer with 27 bp flanks. C) as A) but with a short dsDNA target encompassing only the R44 protospacer. A-B) Cascade concentration range is 625, 125, 25, 5 and 1 ng/µl. D) Effect of the Protospacer Adjacent Motif (PAM) on target dsDNA recognition. Cascade concentration range is 160, 80, 40, 20, 10 and 5 ng/µl.

It appears that Cascade recognizes dsDNA targets by base pairing of the crRNA spacer sequence with the complementary DNA strand. Analysis of long dsDNA targets (protospacer with 27 bp flanks) showed that both strands shifted (Fig. 3B) due to base pairing of the flanking regions. Only the complementary strand shifted when short dsDNA targets (corresponding to the protospacer) were used, indicating that crRNA base pairing with the complementary strand caused displacement of the non-complementary strand (Fig. 3C). The recognition of the target was independent of the recently identified protospacer adjacent motif (PAM) (Fig. 3D), a conserved nucleotide sequence located near one end of the crRNAtargeted region (Mojica *et al*., 2009). Cascade-mediated target DNA cleavage was not observed under any of the conditions tested. Binding to complementary ssRNA could be demonstrated (Fig. 2I), but this did not result in target RNA cleavage. In contrast to dsDNA, no binding to dsRNA was observed (Fig. 2J). In addition, very little non-specific interaction occurred between Cascade and ss- or dsRNA (Figs. 2I and 2J).

Role of the protospacer adjacent motif

While PAMs apparently do not influence recognition of the protospacer by Cascade in vitro, we investigated whether the motif plays a role in the recognition and interference of targeted regions in vivo. For this we constructed five different CRISPRs (J1-5) each containing one spacer targeting nearly the same protospacer region on the phage Lambda (λ) genome (coordinates: 18066-18101). A protospacer region was chosen on the template (noncoding) DNA strand of the gene encoding the tail protein (J) of the phage. Of the five CRISPRs, only the J3 CRISPR targeting a protospacer region directly upstream of the *E. coli* K12-specific PAM sequence 5'-protospacer-CNT-3' (Mojica *et al*., 2009) was effective. This CRISPR reduced host sensitivity for phage λ by approximately 10^3 fold (Fig. 4). Four other CRISPRs (J1, J2, J4, J5) targeting a protospacer region shifted either one or two nucleotides upstream or downstream in the phage genome were completely ineffective (Fig. 4).

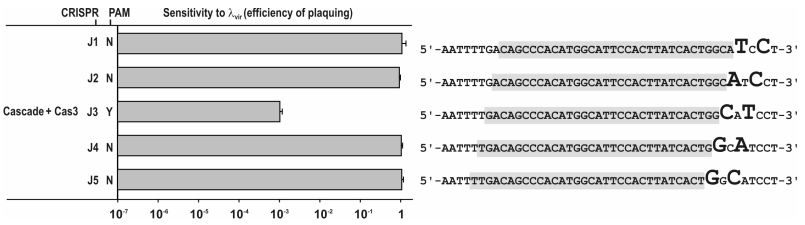


Figure 4. Effect of the protospacer adjacent motif (PAM) on phage resistance. A histogram is shown of the efficiency of plaquing (EOP) in the presence of Cascade, Cas3 and one of five CRISPRs (J1-J5). The protospacer sequence and directly adjacent nucleotides are shown

Subunit stoichiometry of Cascade

To understand the structural basis of the interaction between Cascade and target DNA, the composition of the Cascade protein assembly was determined using an array of mass spectrometric analyses. Denaturing and tandem mass spectrometry analyses resulted in accurate mass measurements for each component of Cascade (not shown). The measured masses CasA, CasB and CasD were in agreement with the expected values and the mass of CasC and CasE with the primary amino acid sequence lacking their N-terminal methionine. The complex would have a mass of 184 kDa when each component would be present at unit stoichiometry. However, analysis of the intact assembly by native mass spectrometry (Heck, 2008) showed two major charge state distributions, corresponding to masses of 405,365 ± 135 Da and 349,399 ± 84 Da (Fig. 5A). A third low intensity charge state distribution around 2,200 m/z was observed with a mass of 42,524 ± 8 Da, which is close to the theoretical molecular weight of a CasB dimer (42,521 Da). Proteolytic removal of the affinity tag on CasB unambiguously confirmed the presence of two CasB copies.

The two major complexes of 405 and 349 kDa likely resemble the intact Cascade and a Cascade sub-complex lacking CasA, in agreement with the mass difference of 55,966 Da. The presence of this sub-complex suggests that at least one CasA copy is located at the periphery of Cascade with a rather low affinity. Tandem mass spectrometry experiments on intact Cascade ions revealed that CasA was the first subunit to be expelled from the complex under collisional activation conditions, again indicative of a peripheral position of CasA within the complex (not shown). In addition to the elimination of CasA, also the loss of CasC was observed by tandem mass spectrometry. Similarly, selection and activation of the 349 kDa Cascade sub-complex showed the loss of one CasD and up to two CasC subunits. Unfortunately, Cascade could not be further disrupted by tandem mass spectrometry, hampering the full assignment of its stoichiometry. Therefore a number of alternative strategies were devised.

Since Cascade was loaded with a single type of crRNA (R44, not shown), the number of bound crRNA molecules could be determined by adding a complementary ssDNA probe. The total molecular weight of Cascade increased by the mass of a single ssDNA-probe, indicating the presence of one crRNA.

For further characterization we used a combined approach of in solution and in gasphase dissociation of Cascade (Lorenzen *et al.*, 2007; Zhou *et al.*, 2008). By adding a low percentage of 2-propanol, Cascade was partially disrupted, resulting in a variety of Cascade sub-complexes (Fig. 5B). In addition to the intact Cascade (405 kDa), we also detected Cascade lacking CasA (349 kDa), and seven additional sub-complexes, the largest

of which is a 307 kDa species. The difference between the 349 kDa and 307 kDa sub-complexes is 42,442 Da and likely reflects the loss of the CasB dimer, which was previously observed (Fig. 5A).

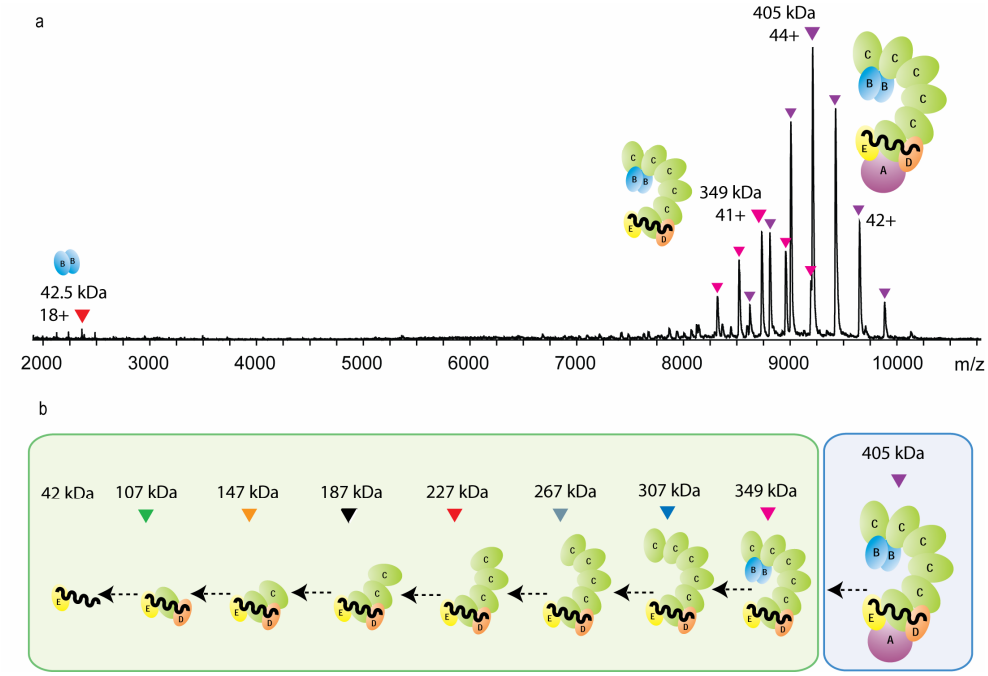


Figure 5. A) Native nano-ESI mass spectrum of Cascade. Two charge state distributions are present at high m/z values, corresponding to complexes of 405 kDa (purple) and 349 kDa (pink). The charge state distribution indicated in red indicates the CasB dimer. B) Cascade (sub)complexes analyzed by native mass spectrometry. The sub-complexes were formed in solution after adding 5% 2-propanol to the buffer solution containing Cascade.

The apparent consecutive loss of five times a mass of around 40 kDa (from 307 kDa down to 107 kDa) is directly evident from this catalog of sub-complexes (Fig. 5B). This implies that at least five CasC subunits are present in intact Cascade. This hypothesis was confirmed by subjecting each of these sub-complexes to tandem mass spectrometry. In addition to CasD, CasE and one crRNA, this analysis revealed the presence of even a sixth CasC subunit in the 107 kDa sub-complex (Fig. 5B). Combining all mass spectrometry data resulted in a Cascade stoichiometry of CasABCDE/crRNA of 1:2:6:1:1/1. The theoretical mass of this complex is in excellent agreement with the experimental mass (405,095 Da versus 405,365 ± 135 Da), band intensities on protein gels (Fig. 1B), and elution profiles on a calibrated size exclusion column (not shown). Similar analyses of purified CasBCDE and CasCDE sub-complexes revealed a mass of 349 and 324 kDa,

respectively, and showed that their stoichiometry were consistent with Cascade. The constituency of the complexes and the elimination of specific subunits in solution and under tandem mass spectrometry conditions allowed for the composition of a topology model for Cascade, which reinforced the assignments of subunits in the structural model presented below.

Structure of Cascade

First insights into the structural organization of Cascade were obtained by single particle electron microscopy (EM) and small angle X-ray scattering (SAXS). The Cascade complex is an elongated particle with no discernable symmetry. The particle has approximate dimensions of 10 x 20 nm, and resembles a seahorse with a curled-up tail (Figs. 6A-C and 7B). The particle displays a striking indentation on one side which gives rise to the head and neck features of the seahorse. Three types of projection maps with minor variation were obtained for Cascade (Figs. 6A-C), indicative of a strong adsorption orientation preference on the carbon support film. In addition to DNA-free Cascade, we also examined the structure of Cascade with target ssDNA bound. This resulted in six groups of projections with substantial differences to one another (Figs. 6D-I). The comparison between DNA-free and DNA-bound Cascade reveals differences in the torso of the seahorse-shaped morphology of Cascade (Fig. 6A vs. 6F), which suggests that Cascade undergoes conformational changes upon target DNA binding.

Subunit localization

Given the availability of stable sub-complexes of Cascade and knowing their subunit compositions allowed for the investigation of subunit localization within the Cascade structure. Some of the projections show a regularly shaped and evenly spaced feature with sharp edges that spans the torso of the complex (Fig. 6I). This regularly-arranged spine structure is interpreted as a hexameric CasC backbone. The position of CasA was determined from difference maps between target DNA bound forms of Cascade and Cascade lacking CasA (CasBCDE) (Figs. 6J-L). The analysis revealed that CasA is located in the curled-up tail of the seahorse (Fig. 6L). The position of CasB was identified using difference maps between target DNA bound CasBCDE and CasCDE complexes (Figs. 6M-O). Apart from the six better resolved CasC subunits in these projections of CasBCDE and CasCDE (Figs. 6M and 6N), the difference map showed one region compatible with two CasB copies, consistent with the CasB homodimer observed with mass spectrometry (Fig.

5A). The CasB dimer is contributing the nose to the seahorse-shaped morphology of Cascade.

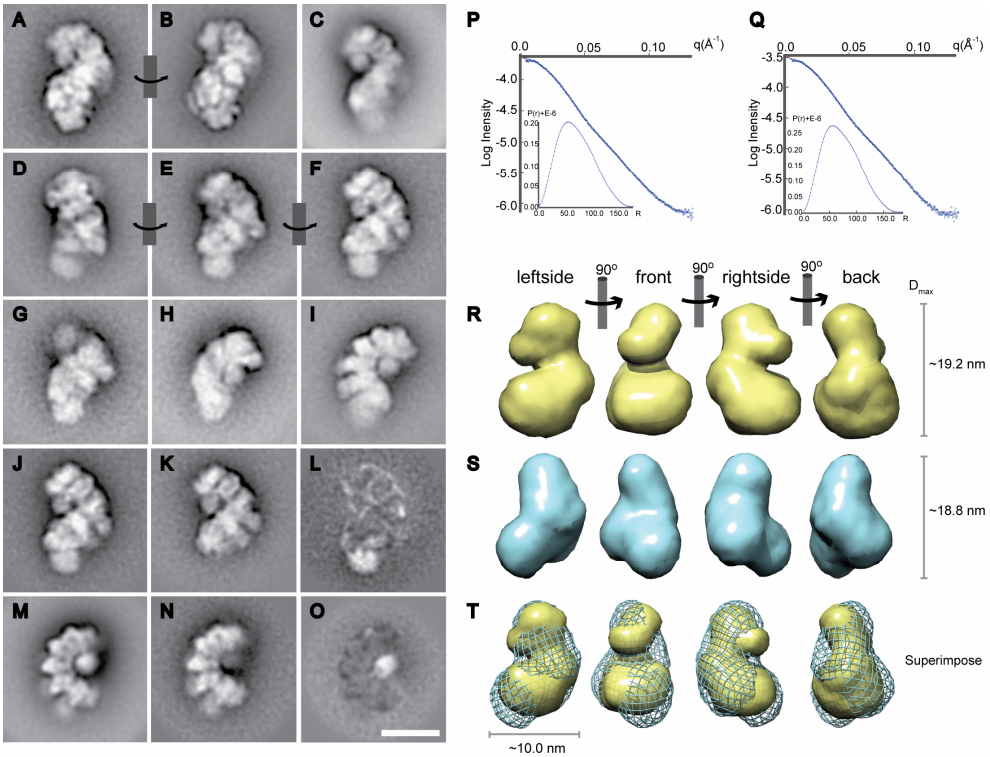


Figure 6. EM and SAXS structure of Cascade. A-C) Cascade projections showing an elongated particle with 20 x 10 nm dimensions and resembles a seahorse. D-I) Cascade projections bound to target ssDNA. J-L) CasA difference map (L) generated from Cascade (J) and CasBCDE (K) projection maps with target ssDNA bound. M-O) CasB difference map (O) generated from CasBCDE (M) and CasCDE (N) projection maps with target ssDNA bound. On average, each image is the sum of 1,500 single particle projections. The scale bar equals 10 nm. P-T) Solution scattering model of Cascade obtained with SAXS. Scattering data for Cascade were collected at of 10 keV (1.24 Å λ) from two protein concentrations, and include scattering vectors (*q*), ranging from (P) 0.015 Å-1 to 0.127 Å-1 for Cascade and (Q) 0.015 $Å^{-1}$ to 0.133 $Å^{-1}$ for Cascade bound to target DNA. The pair-distribution function (insert) indicates that the radius of gyration for both particles is ~5.6 nm. R) The Cascade particle has a maximum dimension, D_{max} ~19.2 nm. S) Cascade changes shape upon binding target DNA from a seahorse to a more boot morphology with a maximum dimension of ~18.8 nm. T) Superposition of the solution structures of Cascade without (yellow) and with target DNA (mesh). Images have been rendered using Chimera (Goddard *et al.*, 2005).

In addition to the position of some of the subunits in the Cascade structure, we also investigated the topological constraints of Cascade. Protein pulldown experiments between CasC and one of the remaining Cascade components showed that CasB, CasD and CasE form stable two-component complexes with CasC, whereas CasA does not (not

shown). The association of this subunit with the CasC backbone is mediated by CasE (not shown).

Solution structure of Cascade

Ab initio small-angle X-ray scattering (SAXS)-based reconstructions of Cascade and Cascade bound to target DNA (Fig. 6) result in well-supported models that are of similar shape and size to what was observed by EM (Fig. 6A vs. 6R). Scattering curves for Cascade (Fig. 6P) and Cascade bound to target DNA (Fig. 6Q) were generated, resulting in a Guinier-based estimations of the radius of gyration (R(g)) of 5.6 nm for both Cascade and Cascade bound to target DNA. Real-space information about the Cascade molecule was obtained by analyzing the pair-distribution function. This analysis shows that the most frequently sampled interatomic distance is 5.61 nm and 5.60 nm for each of the two complexes, which is nearly identical to that estimated by the Guinier analysis.

Ten independent reconstructions of the two samples were in good agreement with one another and were used to generate the average models. The asymmetric seahorse-shaped morphology of Cascade is also evident in the SAXS model (Fig. 6R). This model highlights a twist in the spine of the particle, which we interpret to be asymmetric stacking of the six CasC subunits. As observed in the EM analysis, the shape of Cascade changes upon binding target DNA, resulting in a shape with fewer prominent features. The most obvious change is loss of the nose-shaped feature and a less compact tail resulting in a more boot-shaped particle (Fig. 6S). Superimposition of the two types of solution structures shows the differences between the DNA-free and DNA-bound state, indicative of conformational changes (Fig. 6T).

Discussion

Composition of Cascade

Cascade is a ribonucleoprotein complex in *E. coli* that plays a central role in CRISPR-based defense against mobile genetic elements, such as phages and conjugative plasmids. Despite its crucial role as the main effector complex, very little is known about how the host counter attack is accomplished. Cascade has a mass of 405 kDa and consists of five proteins (CasA, CasB, CasC, CasD and CasE) and one guide RNA (Brouns *et al*., 2008). The core of the protein complex consists of six CasC (Cse4) subunits, presumably to provide a structural backbone for the other subunits and the crRNA. Given the presence of CasC-homologs (COG1857) in the majority Cas subtypes (Cse4, Csd2, Csh2, Cst2, Csa2, Csy2)

(Makarova *et al*., 2006), these proteins could be components of Cascade-like protein complexes from other microbes. The next most frequently found subunit of the complex is CasD (Cas5e), which belongs to the Cas5-type (Cas5e, Cas5d, Cas5h, Cas5t, Cas5a, Csy3) RAMP protein class (Repeat-Associated Mysterious Proteins). This protein co-occurs with CasC-homologs in five subtypes (Makarova *et al*., 2006). CasD is present in a single copy in Cascade and together with six CasC subunits and one CasE (Cse3) subunit part of a minimal Cascade core that accommodates a single mature crRNA. This minimal core is expanded by a dimer of CasB (Cse2), a positively charged protein (pI 9.2). A crystal structure of the CasB from *Thermus thermophilus* HB8 shows a strictly α-helical protein with a conserved basic surface patch that suggests a role in nucleic acid interactions (Agari *et al*., 2008) Cascade is completed by one copy of CasA (Cse1), the largest and most loosely attached subunit of the five. Gene synteny analysis of the various Cas gene clusters shows that CasA and CasB only occur in the Cse subtype, and that the CasAB pair is substituted by other Cas proteins (e.g. Csy1) containing the CasCD core (van der Oost *et al*., 2009).

The crRNA is remarkably stable when bound by Cascade or the CasBCDE and CasCDE sub-complexes, indicating that it is tightly associated with the CasCDE core of the protein complex and well shielded from cellular ribonuclease activities. Yet it is sufficiently exposed to allow for base pairing with complementary nucleic acids. Interestingly, the length of the 5' handle is conserved among crRNAs from *E. coli*, *S. epidermidis* and psiRNAs from *P. furiosus* (Brouns *et al*., 2008; Carte *et al*., 2008; Marraffini *et al*., 2008), suggesting a general mechanism of binding. The presence of a 5'-hydroxyl group is also observed in Cas6-generated crRNA (Carte et al., 2008), but not in eukaryotic small interfering RNA and microRNA (miRNA) which need to be 5'-phosphorylated in order to bind to Argonaute and serve as a guide for the RNA-induced silencing complex (RISC) (Ma *et al*., 2005). *E. coli* crRNAs are unmodified, unlike some plant miRNAs and vertebrate piwi-interacting RNAs, which are methylated at the ribose of the last nucleotide, possibly to enhance stability (Houwing *et al*., 2007; Yu *et al*., 2005). The cyclic 2',3'-phosphate of crRNA results from a metal ion-independent endonuclease activity on the 3'-side of the phosphodiester bond of pre-crRNA, as suggested previously (Carte *et al*., 2008). These initial pre-crRNA endonuclease cleavage products are the mature form of the crRNA in *E. coli*, in contrast to crRNAs from *P. furiosus*, *S. solfataricus* and *S. epidermidis* which are trimmed at the 3'-end (Hale *et al*., 2009; Lillestøl *et al*., 2009; Marraffini *et al*., 2010).

Cascade targets DNA

Multiple lines of evidence suggest that crRNAs directly target invader DNA in the *E. coli* (Cse) and *S. epidermidis* (Csm) model systems (Brouns *et al*., 2008; Marraffini *et al*., 2008). However, molecular evidence of Cas-effector complexes recognizing their target DNA has been lacking, limiting our understanding of how CRISPR/Cas systems operate. Binding studies showed that Cascade is guided by the crRNA towards dsDNA sequences containing the protospacer without additional co-factors such as ATP. This surprising characteristic of Cascade is physiologically relevant, since most cellular and invader DNA is of double-stranded nature. The ATP-independence of this scanning process allows cells to continuously survey nucleic acids for crRNA matches without major energy investments. In addition to sequence-specific DNA recognition, Cascade also interacts non-specifically with DNA. Cascade sub-complexes lacking CasA, however, display only sequence-specific DNA recognition. Direct competition assays between Cascade and CasBCDE demonstrate the superiority of Cascade in locating the protospacer DNA, suggesting a role for CasA as an enhancer. Cascade thus appears to interact with DNA in two different modes, non-specifically via CasA and sequence-specifically via Watson-Crick base pairing of the crRNA (Fig. 7A). We hypothesize that the non-specific affinity of Cascade for DNA enables Cascade to associate fast enough with incoming foreign DNA to neutralize a rapidly proliferating phage or conjugative plasmid in a cell. Further analysis showed that although Cascade exhibits little non-specific affinity for RNA, it binds complementary ssRNA as well. Base pairing of the crRNA with mRNA occurs when the coding strand of an invading nucleic acid is targeted. This could titrate out unpaired Cascade-bound crRNAs, reducing the effectiveness of the immune response (Brouns *et al*., 2008). Cascade does not recognize a protospacer within a dsRNA molecule, indicating that the ability to locate protospacers in a nucleic acid duplex is restricted to dsDNA. Recently, a distinct Cas protein complex of the Cmr-subtype from *P. furiosus* was shown to cleave ssRNA complementary to the bound guide RNA (called psiRNA, Hale *et al*., 2009), akin to RISC in eukaryotes. Target RNA hydrolysis takes place 14 bases from the 3'-end of the psiRNA, suggesting that target RNA cleavage occurs by a ruler mechanism. The difference in target nucleic acid - RNA or DNA - between the Cmr and Cse-subtypes illustrates the remarkable diversity present among CRISPR/Cas systems.

Requirements for CRISPR-based defense

Despite the fact that stable sub-complexes of Cascade can be produced, which cleave and retain crRNA and recognize their target DNA *in vitro*, it is evident that all five protein

subunits of Cascade, a targeting crRNA and the predicted nuclease/helicase Cas3 are required for antiviral defense *in vivo*. In addition, viral plaque assays showed that the protospacer adjacent motif (PAM) (Mojica *et al*., 2009) or CRISPR motif (Deveau *et al*., 2008; Horvath *et al*., 2008) plays an important role at the interference stage as well. The relevance of these short nucleotide sequences was originally shown in *S. thermophilus* by sequencing phages that had overcome host immunity by mutating a single nucleotide of the motif (Deveau *et al*., 2008). The emerging picture of the motif is one of ubiquitous occurrence, variation in sequence and size (2-4 nt), as well as placement on either side of the protospacer (Lillestøl *et al*., 2009; Mojica *et al*., 2009; Semenova *et al*., 2009). Despite the fact that Cascade recognizes protospacers with and without PAM equally well, only a protospacer flanked by a PAM is subject to interference in vivo. This suggests that the interference is not just due to binding of Cascade to the target DNA - blocking transcription or replication - but to subsequent interference events. We speculate these events include target DNA cleavage by the HD-domain of Cas3, since deoxyribonuclease activity has been demonstrated for the standalone HD-domain from *Sulfolobus solfataricus* (Han *et al*., 2009).

A recent study in *S. epidermidis* showed that differential complementarity of the 5'-handle of the crRNA with the downstream protospacer flank allows discrimination between self DNA (the CRISPR) and non-self DNA (the target) (Marraffini *et al*., 2010). While non-complementarity of the 5'-handle results in a sequence that is targeted, base pairing of at least nucleotide 5, 6 and 7 of the 5'-handle provides protection. Cas protein proofreading of paired vs. unpaired structures in this region was speculated to either abort or enable downstream interference processes (Marraffini *et al*., 2010). Although no PAM is known for the *S. epidermidis* repeat type, PAMs for other repeat types are often present in this apparently critical region. In *E. coli*, the PAM sequence (i.e. 5'-protospacer-CNT-3') gives rise to non-basepaired rC:dT and a basepaired rG:dC between the sixth and eighth base of the 5'-handle (Fig. 7A). This signature in the target DNA-Cascade complex might serve to verify that a target sequence has been located before downstream interference processes are commenced.

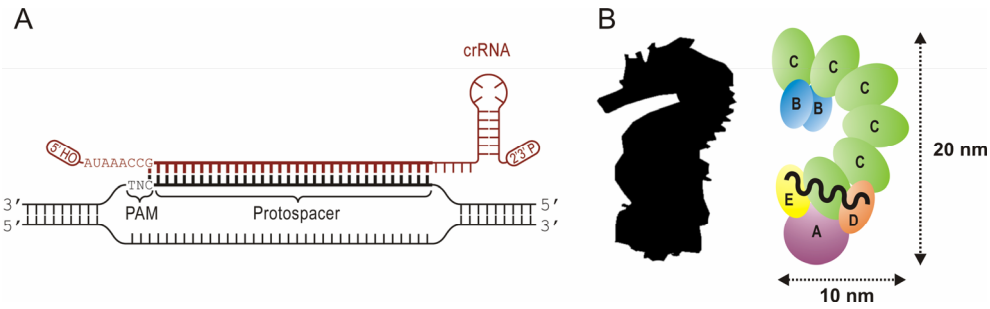


Figure 7. A) Schematic diagram of crRNA base paired to double stranded target DNA, indicating the local strand displacement, and the additional rG-dC basepair between the eighth base of the crRNA (rG) with the PAM (dC). B) Seahorse morphology and structural model of Cascade. Note that the locations of CasD, CasE and crRNA in the structure are ambiguous.

Structural basis for target DNA recognition

Cascade displays an unusual seahorse-shaped architecture. The backbone of this structure is composed of six CasC subunits that are arranged in an arch, representing the torso of the Cascade particle. CasC only assembles into this stable hexameric arrangement in the presence of CasD, CasE and the CasE-generated crRNA. Although the location of these latter three components in the structure could not be determined by direct methods, mass spectrometry showed the existence of a 107 kDa 2-propanol-induced sub-complex containing single copies of CasC, CasD, CasE and crRNA (Fig. 5B). This subassembly of Cascade implies that these four components are in close proximity of each other. Further disassembly of Cascade reveals a CasE-crRNA particle, suggesting that CasE holds on tightly to the crRNA after pre-crRNA cleavage. Despite the fact that the location of the crRNA remains unclear, we expect that the 61 nucleotide molecule is bound by a substantial part of Cascade. In an unpaired state, the crRNA likely experiences some conformational flexibility, whereas in a DNA base paired state, the conformation of the spacer sequence is constrained by the formation of the RNA:DNA helix, which is commonly A-form. This constraint might induce the conformational change observed in the EM and SAXS structures of Cascade, and may serve to recruit Cas3, which is required for neutralizing the targeted invader. The length of the 32 bp spacer sequence base paired to its target ssDNA is approximately 8.2 nm. Together with the 5' and 3' handles, the crRNA could be up to 10 nm in length. These combined observations result in a structural model in which the position of CasA, CasB and CasC can be assigned with confidence, contributing the curled-up tail, nose and torso of the seahorse, respectively (Fig. 7B). The model predicts that CasD, CasE and at least part of the crRNA are located in the main

body of Cascade at the tail-end of the CasC spine, in the proximity of CasA. The crRNA could also be bound along the spine of the CasC backbone and in this way define the number of CasC subunits by the length of the crRNA.

In conclusion, Cascade is a highly unusual ribonucleoprotein complex capable of specifically recognizing double stranded invader DNA archived in the CRISPR blacklist. The direct attack on invader DNA rather than its RNA transcripts may be a defense strategy that effectively neutralizes the source of the invasion. Even though many questions remain, we provide first insights into the unique mechanism and structure of the CRISPR defense machinery that enables prokaryotes to combat selfish mobile elements.

Experimental procedures

Protein production and purification

Cascade complexes were produced and purified as described previously (Brouns *et al*., 2008), using the expression plasmids listed in Table S2. Cascade was routinely purified with an N-terminal Strep-tag II fused to CasB (or CasC in CasCDE). Size exclusion chromatography (Superdex 200 HR 10/30 (GE)) was performed using 20 mM Tris-HCl (pH 8.0), 0.1 M NaCl, 1 mM dithiotreitol. Cascade preparations (~0.3 mg) were incubated with DNase I (Invitrogen) in the presence of 2.5 mM $MgCl_2$ for 15 min at 37 °C prior to size exclusion analysis. Copurified nucleic acids were isolated by extraction using an equal volume of phenol:chloroform:isoamylalcohol (25:24:1) pH 8.0 (Fluka), and incubated with either DNase I (Invitrogen) supplemented with 2.5 mM $MgCl_2$ or RNase A (Fermentas) for 10 min at 37 ºC.

Lambda phage infection

Plaque assays were performed using bacteriophage Lambda and the efficiency of plaguing (EOP) was calculated as described previously (Brouns *et al*., 2008)

HPLC purification of crRNA

All samples were analyzed by ion-pair reversed-phased-HPLC on an Agilent 1100 HPLC with UV_{260nm} detector (Agilent) using a DNAsep column 50 mm × 4.6 mm I. D. (Transgenomic, San Jose, CA). The chromatographic analysis was performed using the following buffer conditions: A) 0.1 M triethylammonium acetate (TEAA) (pH 7.0) (Fluka); B) buffer A with 25% LC MS grade acetonitrile (v/v) (Fisher). The crRNA was obtained by

injecting purified intact Cascade at 75 °C using a linear gradient starting at 15% buffer B and extending to 60% B in 12.5 min, followed by a linear extension to 100% B over 2 min at a flow rate of 1.0 ml/min. Hydrolysis of the cyclic phosphate terminus was performed by incubating the HPLC-purified crRNA in a final concentration of 0.1 M HCl at 4 ºC for 1 hour. The samples were concentrated to 5-10 µl on a vacuum concentrator (Eppendorf) prior to ESI-MS analysis.

ESI-MS analysis of crRNA

Electrospray Ionization Mass spectrometry was performed in negative mode using an UHR-TOF mass spectrometer (maXis) or an HCT Ultra PTM Discovery instrument (both Bruker Daltonics), coupled to an online capillary liquid chromatography system (Ultimate 3000, Dionex, UK). RNA separations were performed using a monolithic (PS-DVB) capillary column (200 µm × 50 µm I.D., Dionex, UK). The chromatography was performed using the following buffer conditions: C) 0.4 M 1,1,1,3,3,3,-Hexafluoro-2-propanol (HFIP, Sigma- Aldrich) adjusted with triethylamine (TEA) to pH 7.0 and 0.1 mM TEAA, and D) buffer C with 50% methanol (v/v) (Fisher). RNA analysis was performed at 50 ºC with 20% buffer D, extending to 40% D in 5 min followed by a linear extension to 60% D over 8 min at a flow rate of 2 µl/min.

Protein and Native Mass spectrometry

Cascade was analyzed in 0.15 M ammonium acetate (pH 8.0) at a protein concentration of 5 µM. This protein preparation was obtained by five sequential concentration and dilution steps at 4 ºC using a centrifugal filter with a cut-off of 10 kDa (Millipore). Proteins were sprayed from borosilicate glass capillaries and analyzed on a LCT electrospray time-of-flight or modified quadrupole time-of-flight instruments (both Waters, UK) adjusted for optimal performance in high mass detection (Tahallah *et al*., 2001; van den Heuvel *et al*., 2006). Exact mass measurements of the individual Cas proteins were acquired under denaturing conditions (50% acetonitrile, 50% MQ, 0.1% formic acid). Sub-complexes in solution were generated by the addition of 2-propanol to the spray solution to a final concentration of 5% (v/v). Instrument settings were as follows; needle voltage ~1.2 kV, cone voltage ~175 V, source pressure 9 mbar. Xenon was used as the collision gas for tandem mass spectrometric analysis at a pressure of 1.5 10^{-2} mbar. The collision voltage varied between 10-200 V.

Electrophoretic mobility shift assays

EMSA was performed by incubating Cascade, CasBCDE or CasCDE with 1 nM ^{32}P-labeled nucleic acid in a binding buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl. Salmon sperm DNA (Invitrogen) competitor concentration was 0.5 mg/ml, unless stated otherwise.

The EMSA reactions were incubated at 37° C for 20-30 min prior to electrophoresis on 5% polyacrylamide gels. The gels were subsequently dried and the radioactivity visualized using phosphor imaging.

DNA targets were gel-purified long oligonucleotides (Isogen Life Sciences or Biolegio) listed in Table S2. The DNA oligonucleotides were end-labeled using γ ^{32}P-ATP (PerkinElmer) and T4 kinase (Fermentas). Double-stranded DNA targets were prepared by annealing complementary oligonucleotides. Remaining ssDNA was removed by Exonuclease I treatment (Fermentas). RNA targets were produced using T7 Maxiscript or T7 Mega Shortscript *in vitro* transcription kits (Ambion) with α ^{32}P-CTP (PerkinElmer). Double stranded RNA targets were prepared by annealing complementary RNA and digesting surplus ssRNA with RNase T1 (Fermentas), followed by phenol extraction.

Electron microscopy

Purified protein samples were negatively stained with 2% uranyl acetate on glow-discharged carbon-coated copper grids. R44 target ssDNA (BG3028, Table S2) was added to Cascade sub-complexes loaded with R44 crRNA in a two-fold excess at least 5 min prior to sample preparation. Electron microscopy was performed on a Philips CM120 equipped with a LaB6 tip operating at 120 kV. Images were recorded with a Gatan 4000 SP 4K slow-scan CCD camera at 130,000× magnification at a pixel size (after binning the images) of 0.23 nm at the specimen level with GRACE software for semi-automated specimen selection and data acquisition (Oostergetel *et al*., 1998). Single particle projections were selected from micrographs mainly by reference-based automated particle selection procedure incorporated into GRIP (GRoningen Image Processing) software (van Heel *et al*., 2000). Approximately 400,000 single particles were selected and extracted from 17,000 electron micrographs. Single particle data sets were analyzed with the GRIP software using multi-reference alignments and no-reference alignments, multivariate statistical analysis, and hierarchical ascendant classification. The final two-dimensional projection maps were calculated from the best resolved classes by summing the best 5–20% of the projections based on the correlation coefficient determined in the alignment step.

Small Angle X-ray Scattering

SAXS data were collected at the Advanced Light Source (Lawrence Berkeley National Laboratory) on beamline 7.3.3. Solution scattering of Cascade was collected at room temperature (~22 °C) using at least two different concentrations (between 1 mg/ml and 18 mg/ml) in a 20 µl sample cell with an X-ray flux of 10 keV (1.24 Å λ). The sample-to-detector-distance was set to 3056.69 m resulting in scattering vectors (q) ranging from 0.012 $Å^{-1}$ to 0.127 $Å^{-1}$ for Cascade, and 0.0147 $Å^{-1}$ to 1.334 $Å^{-1}$ for Cascade bound to target DNA. Scatter plots for the low and high concentrations were merged and background subtracted using PRIMUS (Konarev *et al*., 2003). One-dimensional scatter curves were transformed and distance distribution functions P(r) were calculated using GNOM (Svergun, 1992). The pairdistribution function P(r), the frequency of interatomic vector lengths within the scattering particle, was calculated from the entire scattering curve I(q), by an indirect Fourier transform using GNOM. Ten independent models of each complex were generated using a simulated annealing method in DAMMIF (Franke *et al*., 2009). *Ab initio* reconstructions for each complex were aligned, filtered and averaged based on occupancy using DAMAVER (Volkov *et al*., 2003). The SAXS bead models were converted to volumetric format using the pdb2vol convolution kernel in the Situs software package (Wriggers *et al*., 1999).

Acknowledgements

We thank L. van Heereveld and Man H. Lai for experimental contributions, and Eric Schaible and Peter Zwart for technical support and for assistance with post processing of SAXS data. This work was financially supported by an NWO Vici grant to JvdO (865.05.001), Veni grants to SJJB (863.08.014) and EvD (700.58.402), NWO TOP grant to EJB, EPSRC and BBSRC grants to MD. ML was financially supported by the Wenner-Gren Foundations, ERW by Spinoza resources awarded to W.M. de Vos. APLS is an RCUK Academic Fellow. BW is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. We thank the Netherlands Proteomics Center for financial support.

References

References

Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E. (1994) Structure at 2.8-Angstrom Resolution of F_1-Atpase from Bovine Heart-Mitochondria. *Nature*, **370**, 621-628

Acín-Pérez, R., Bayona-Bafaluy, M.P., Fernández-Silva, P., Moreno-Loshuertos, R., Pérez-Martos, A., Bruno, C., Moraes, C.T., and Enríquez, J.A. (2004) Respiratory complex III is required to maintain complex I in mammalian mitochondria. *Mol. Cell*, **13**, 805-815

Acín-Pérez, R., Fernández-Silva, P., Peleato, M.L., Pérez-Martos, A., and Enríquez, J.A. (2008) Respiratory Active Mitochondrial Supercomplexes. *Mol. Cell*, **32**, 529-539

Agari, Y., Yokoyama, S., Kuramitsu, S., and Shinkai, A. (2008) X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermaphilus* HB8. *Proteins*, **73**, 1063-1067

Allen, J.F. and Martin, W. (2007) Evolutionary biology - Out of thin air. *Nature*, **445**, 610-612

Allen, R.D., Schroeder, C.C., and Fok, A.K. (1989) An Investigation of Mitochondrial Inner Membranes by Rapid-Freeze Deep-Etch Techniques. *J. Cell Biol.*, **108**, 2233-2240

Anderson, S.L. and McIntosh, L. (1991) Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: a blue-light-requiring process. *J. Bacteriol*, **173**, 2761-2767

Aravind, L. and Koonin, E.V. (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sc.*, **23**, 468-472

Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R.A., and Schägger, H. (1998) Yeast mitochondrial F_1F_0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. *EMBO J.*, **17**, 7170-7178

Aseeva, E., Ossenbuhl, F., Eichacker, L.A., Wanner, G., Soll, J., and Vothknecht, U.C. (2004) Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. *J. Biol. Chem.*, **279**, 35535-35541

Aseeva, E., Ossenbuhl, F., Sippel, C., Cho, W.K., Stein, B., Eichacker, L.A., Meurer, J., Wanner, G., Westhoff, P., Soll, J., and Vothknecht, U.C. (2007) Vipp1 is required for basic thylakoid membrane formation but not for the assembly of thylakoid protein complexes. *Plant Physiol. Biochem.*, **45**, 119-128

Baranova, E.A., Holt, P.J., and Sazanov, L.A. (2007) Projection structure of the membrane domain of *Escherichia coli* respiratory complex I at 8 angstrom resolution. *J. Mol. Biol.*, **366**, 140-154

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712

Bekker, A., Holland, H.D., Wang, P.L., Rumble, D., Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004) Dating the rise of atmospheric oxygen. *Nature*, **427**, 117-120

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A., and Yakunin, A.F. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.*, **283**, 20361-20371

Bianchi, C., Genova, M.L., Castelli, G.P., and Lenaz, G. (2004) The mitochondrial respiratory chain is partially organized in a supercomplex assembly - Kinetic evidence using flux control analysis. *J. Biol. Chem.*, **279**, 36562-36569

Boekema, E.J. (1991) Negative Staining of Integral Membrane-Proteins. *Micron and Microscopica Acta*, **22**, 361-369

Boekema, E.J. and Braun, H.P. (2007) Supramolecular structure of the mitochondrial oxidative phosphorylation system. *J. Biol. Chem.*, **282**, 1-4

Boekema, E.J., Folea, M., and Kouril, R. (2009) Single particle electron microscopy. *Photosyn. Res.*, **102**, 189-196

Boekema, E.J., Hifney, A., Yakushevska, A.E., Piotrowski, M., Keegstra, W., Berry, S., Michel, K.P., Pistorius, E.K., and Kruip, J. (2001) A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature*, **412**, 745-748

Bolotin, A., Ouinquis, B., Sorokin, A., and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiol.*, **151**, 2551-2561

Boumans, H., Grivell, L.A., and Berden, J.A. (1998) The respiratory chain in yeast behaves as a single functional unit. *J. Biol. Chem.*, **273**, 4872-4877

Braun, H.P. and Schmitz, U.K. (1992) Affinity Purification of Cytochrome-*C* Reductase from Potato Mitochondria. *Eur. J. Biochem.*, **208**, 761-767

Brissette, J.L., Russel, M., Weiner, L., and Model, P. (1990) Phage shock protein, a stress protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A*, **87**, 862-866

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008) Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, **321**, 960-964

Bryant, D.A. and Frigaard, N.U. (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol.*, **14**, 488-496

Bultema, J.B., Braun, H.P., Boekema, E.J., and Kouril, R. (2009) Megacomplex organization of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato. *Biochim. Biophys. Acta*, **1787**, 60-67

Carte, J., Wang, R.Y., Li, H., Terns, R.M., and Terns, M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, **22**, 3489-3496

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnol. Adv.*, **25**, 294-306

Coffey, A. and Ross, R.P. (2002) Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. *Antonie van Leeuwenhoek*, **82**, 303-321

Cole, C., Barber, J.D., and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucl.Acids Res.*, **36**, W197-W201

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.Sect.D*, **50**, 760-763

Compton, L.A. and Johnson, W.C. (1986) Analysis of Protein Circular-Dichroism Spectra for Secondary Structure Using A Simple Matrix Multiplication. *Anal. Biochem.*, **155**, 155-167

Cournac, L., Guedeney, G., Peltier, G., and Vignais, P.M. (2004) Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. *J. Bacteriol*, **186**, 1737-1746

Daines, D.A. and Silver, R.P. (2000) Evidence for Multimerization of New Proteins Involved in Polysialic Acid Synthesis in *Escherichia co*li K1 Using Improved LexA-Based Vectors. *J. Bacteriol*, **182**, 5267-5270

De Carlo, S., Boisset, N., and Hoenger, A. (2008) High-resolution single-particle 3D analysis on GroEL prepared by cryo-negative staining. *Micron*, **39**, 934-943

De Rosier, D.J. and Klug, A. (1968) Reconstruction of 3 Dimensional Structures from Electron Micrographs. *Nature*, **217**, 130-&

DeLisa, M.P., Lee, P., Palmer, T., and Georgiou, G. (2004) Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. *J. Bacteriol*, **186**, 366-373

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008) Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol*, **190**, 1390-1400

Diaz, F., Fukui, H., Garcia, S., and Moraes, C.T. (2006) Cytochrome *c* oxidase is required for the assembly/stability of respiratory complex I in mouse fibroblasts. *Mol. Cell. Biol.*, **26**, 4872-4881

Dickman, M.J. and Hornby, D.P. (2006) Enrichment and analysis of RNA centered on ion pair reverse phase methodology. *RNA*, **12**, 691-696

Dickson, V.K., Silvester, J.A., Fearnley, I.M., Leslie, A.G.W., and Walker, J.E. (2006) On the structure of the stator of the mitochondrial ATP synthase. *EMBO J.*, **25**, 2911-2918

Dmitrova, M., Younes-Cauet, G., Oertel-Buchheit, P., Porte, D., Schnarr, M., and Granger-Schnarr, M. (1998) A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in *Escherichia coli*. *Mol. Gen. Genet.*, **257**, 205-212

Dudkina, N.V., Eubel, H., Keegstra, W., Boekema, E.J., and Braun, H.P. (2005a) Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. *Proc. Natl. Acad. Sci. U. S. A*, **102**, 3225-3229

Dudkina, N.V., Heinemeyer, J., Keegstra, W., Boekema, E.J., and Braun, H.P. (2005b) Structure of dimeric ATP synthase from mitochondria: An angular association of monomers induces the strong curvature of the inner membrane. *FEBS Letters*, **579**, 5769-5772

Dudkina, N.V., Heinemeyer, J., Sunderhaus, S., Boekema, E.J., and Braun, H.P. (2006a) Respiratory chain supercomplexes in the plant mitochondrial membrane. *Trends Plant Sc.*, **11**, 232-240

Dudkina, N.V., Kouril, R., Peters, K., Braun, H.P., and Boekema, E.J. (2010a) Structure and function of mitochondrial supercomplexes. *Biochim. Biophys. Acta*, In Press, Corrected Proof

Dudkina, N.V., Oostergetel, G.T., Lewejohann, D., Braun, H.P., and Boekema, E.J. (2010b) Row-like organization of ATP synthase in intact mitochondria determined by cryo-electron tomography. *Biochim. Biophys. Acta*, **1797**, 272-277

Dudkina, N.V., Sunderhaus, S., Braun, H.P., and Boekema, E.J. (2006b) Characterization of dimeric ATP synthase and cristae membrane ultrastructure from *Saccharomyces* and *Polytomella* mitochondria. *FEBS Letters*, **580**, 3427-3432

Dühring, U., Irrgang, K.D., Lunser, K., Kehr, J., and Wilde, A. (2006) Analysis of photosynthetic complexes from a cyanobacterial *ycf37* mutant. *Biochim. Biophys. Acta*, **1757**, 3-11

Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S., and Kuramitsu, S. (2006) Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Science*, **15**, 1494-1499

Elderkin, S., Jones, S., Schumacher, J., Studholme, D., and Buck, M. (2002) Mechanism of action of the *Escherichia coli* phage shock protein PspA in repression of the AAA family transcription factor PspF. *J. Mol. Biol.*, **320**, 23-37

Eubel, H., Heinemeyer, J., and Braun, H.P. (2004) Identification and characterization of respirasomes in potato mitochondria. *Plant Phys.*, **134**, 1450-1459

Eubel, H., Jansch, L., and Braun, H.P. (2003) New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. *Plant Phys.*, **133**, 274-286

Faruqi, A.R. and Henderson, R. (2007) Electronic detectors for electron microscopy. *Curr. Opin. Struc. Biol.*, **17**, 549-555

Finger, C., Volkmer, T., Prodohl, A., Otzen, D.E., Engelman, D.M., and Schneider, D. (2006) The stability of transmembrane helix interactions measured in a biological membrane. *J. Mol. Biol.*, **358**, 1221-1228

Frank, J. (2002) Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu. Rev. Biophys. Biomol. Struc.*, **31**, 303-319

Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J.Appl.Cryst.*, **42**, 342-346

Fuhrmann, E., Bultema, J.B., Kahmann, U., Rupprecht, E., Boekema, E.J., and Schneider, D. (2009a) The Vesicle-inducing Protein 1 from *Synechocystis* sp. PCC 6803 Organizes into Diverse Higher-Ordered Ring Structures. *Mol.Biol.Cell*, **20**, 4620-4628

Fuhrmann, E., Gathmann, S., Rupprecht, E., Golecki, J., and Schneider, D. (2009b) Thylakoid membrane reduction affects the photosystem stoichiometry in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol.*, **149**, 735-744

Gao, H. and Xu, X.D. (2009) Depletion of Vipp1 in Synechocystis sp PCC 6803 affects photosynthetic activity before the loss of thylakoid membranes. *FEMS Microbiol. L.*, **292**, 63-70

Gathmann, S., Rupprecht, E., Kahmann, U., and Schneider, D. (2008) A conserved structure and function of the YidC homologous protein slr1471 from *Synechoeystis* sp PCC 6803. *J. Microbiol. Biotechnol.*, **18**, 1090-1094

Gibbons, C., Montgomery, M.G., Leslie, A.G.W., and Walker, J.E. (2000) The structure of the central stalk in bovine F-1-ATPase at 2.4 angstrom resolution. *Nat. Struct. Biol.*, **7**, 1055-1061

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2005) Software extensions to UCSF Chimera for interactive visualization of large molecular assemblies. *Structure*, **13**, 473-482

Godde, J.S. and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.*, **62**, 718-729

Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003) Molecular architecture of the multiprotein splicing factor SF3b. *Science*, **300**, 980-984

Grissa, I., Vergnaud, G., and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *Bmc Bioinformatics*, **8**,

Hackenbrock, C.R., Chazotte, B., and Gupte, S.S. (1986) The Random Collision Model and A Critical-Assessment of Diffusion and Collision in Mitochondrial Electron-Transport. *J. Bioenerg. Biomembr.*, **18**, 331-368

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *Plos Computational Biology*, **1**, 474-483

Hale, C.R., Kleppe, K., Terns, R.M., and Terns, M.P. (2008) Prokaryotic silencing (psi) RNAs in *Pyrococcus furiosus*. *RNA*, **14**, 2572-2579

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009) RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, **139**, 945-956

Han, D. and Krauss, G. (2009) Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Letters*, **583**, 771-776

Hankamer, B., Lehr, F., Rupprecht, J., Mussgnug, J.H., Posten, C., and Kruse, O. (2007) Photosynthetic biomass and H_2 production by green algae: from bioengineering to bioreactor scale-up. *Physiologia Plantarum*, **131**, 10-21

Hankamer, B.D., Elderkin, S.L., Buck, M., and Nield, J. (2004) Organization of the AAA(+) adaptor protein PspA is an oligomeric ring. *J. Biol. Chem.*, **279**, 8862-8866

Heck, A.J.R. (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods*, **5**, 927-933

Heinemeyer, J., Braun, H.P., Boekema, E.J., and Kouril, R. (2007) A structural model of the cytochrome *c* reductase/oxidase supercomplex from yeast mitochondria. *J. Biol. Chem.*, **282**, 12240-12248

Herpen, R.E.M.A., Tjeertes, J.V., Mulders, S.A.M., Ophuis, R.J.A.O., Wieringa, B., and Wansink, D.G. (2006) Coiled-coil interactions modulate multimerization, mitochondrial binding and kinase activity of myotonic dystrophy protein kinase splice isoforms. *FEBS J.*, **273**, 1124-1136

Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*, **327**, 167-170

Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., and Barrangou, R. (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.*, **131**, 62-70

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol*, **190**, 1401-1412

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., Plasterk, R.H.A., Hannon, G.J., Draper, B.W., and Ketting, R.F. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell*, **129**, 69-82

Jansen, R., van Embden, J.D.A., Gaastra, W., and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565-1575

Karginov, F.V. and Hannon, G.J. (2010) The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Mol. Cell*, **37**, 7-19

Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Prot.*, **4**, 363-371

Knight, J.A. (2000) Review: Free radicals, antioxidants, and the immune system. *Ann. Clin. Lab. Sc.*, **30**, 145-158

Kobayashi, R., Suzuki, T., and Yoshida, M. (2007) *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. *Mol. Microbiol.*, **66**, 100-109

Koenig, F. and Schmidt, M. (1995) Gloeobacter Violaceus - Investigation of An Unusual Photosynthetic Apparatus - Absence of the Long-Wavelength Emission of Photosystem-I in 77 K Fluorescence-Spectra. *Physiologia Plantarum*, **94**, 621-628

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J.Appl.Cryst.*, **36**, 1277-1282

Koonin, E.V. and Wolf, Y.I. (2009) Is evolution Darwinian or/and Lamarckian? *Biology Direct*, **4**,

Kouril, R., Yeremenko, N., D'Haene, S., Oostergetel, G.T., Matthijs, H.C.P., Dekker, J.P., and Boekema, E.J. (2005) Supercomplexes of IsiA and photosystem I in a mutant lacking subunit PsaL. *Biochim. Biophys. Acta*, **1706**, 262-266

Kouril, R., Yeremenko, N., D'Haene, S., Yakushevska, A.E., Keegstra, W., Matthijs, H.C.P., Dekker, J.P., and Boekema, E.J. (2003) Photosystem I trimers from *Synechocystis* PCC 6803 lacking the PsaF and PsaJ subunits bind an IsiA ring of 17 units. *Biochim. Biophys. Acta*, **1607**, 1-4

Krause, F., Reifschneider, N.H., Goto, S., and Dencher, N.A. (2005) Active oligomeric ATP synthases in mammalian mitochondria. *Biochem. Biophys. Res. Comm.*, **329**, 583-590

Kroll, D., Meierhoff, K., Bechtold, N., Kinoshita, M., Westphal, S., Vothknecht, U.C., Soll, J., and Westhoff, P. (2001) VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. *Proc. Natl. Acad. Sci. U. S. A*, **98**, 4238-4242

Kunin, V., Sorek, R., and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biology*, **8**,

Laemmli, U.K. (1970) Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature*, **227**, 680-&

Lantz, M., Svensson, M., Björnsson, L., and Börjesson, P. (2007) The prospects for an expansion of biogas systems in Sweden: Incentives, barriers and potentials. *Energy Policy*, **35**, 1830-1843

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948

Lee, S.J., Yamashita, E., Abe, T., Fukumoto, Y., Tsukihara, T., Shinzawa-Itoh, K., Ueda, H., and Yoshikawa, S. (2001) Intermonomer interactions in dimer of bovine heart cytochrome *c* oxidase. *Acta Crystallogr. Sect. D.*, **57**, 941-947

Lee, T.T., Agarwalla, S., and Stroud, R.M. (2004) Crystal structure of RumA, an iron-sulfur cluster containing *E coli* ribosomal RNA 5-methyluridine methyltransferase. *Structure*, **12**, 397-407

Lenaz, G. and Genova, M.L. (2009) Structural and functional organization of the mitochondrial respiratory chain: A dynamic super-assembly. *Int. J. Biochem. Cell Biol.*, **41**, 1750-1772

Li, H.M., Kaneko, Y., and Keegstra, K. (1994) Molecular-Cloning of A Chloroplastic Protein Associated with Both the Envelope and Thylakoid Membranes. *Plant Mol. Biol.*, **25**, 619-632

Li, Y., Horsman, M., Wu, N., Lan, C.Q., and Dubois-Calero, N. (2008) Biofuels from microalgae. *Biotechnol. Prog.*, **24**, 815-820

Liberton, M., Berg, R.H., Heuser, J., Roth, R., and Pakrasi, H.B. (2006) Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp strain PCC 6803. *Protoplasma*, **227**, 129-138

Lillestøl, R.K., Redder, P., Garrett, R.A., and Brügger, K. (2006) A putative viral defence mechanism in archaeal cells. *Archaea*, **2**, 59-72

Lillestøl, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009) CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.*, **72**, 259-272

Liu, C., Willmund, F., Golecki, J.R., Cacace, S., Hess, B., Markert, C., and Schroda, M. (2007) The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in *Chlamydomonas*. *Plant J.*, **50**, 265-277

Liu, C., Willmund, F., Whitelegge, J.P., Hawat, S., Knapp, B., Lodha, M., and Schroda, M. (2005) J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that interacts with vesicle-inducing protein in plastids 1. *Mol.Biol.Cell*, **16**, 1165-1177

Liu, X. and Curtiss, R. (2009) Nickel-inducible lysis system in *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. U. S. A*, **106**, 21550-21554

Lorenzen, K., Vannini, A., Crarner, P., and Heck, A.J.R. (2007) Structural biology of RNA polymerase III: Mass spectrometry elucidates subcomplex architecture. *Structure*, **15**, 1237-1245

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, **434**, 666-670

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, **1**,

Manavalan, P. and Johnson, W.C. (1987) Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular-Dichroism Spectra. *Anal. Biochem.*, **167**, 76-85

Marchetti, J.M., Miguel, V.U., and Errazu, A.F. (2007) Possible methods for biodiesel production. *Renewable Sustainable Energy Rev.*, **11**, 1300-1311

Marques, I., Dencher, N.A., Videira, A., and Krause, F. (2007) Supramolecular organization of the respiratory chain in *Neurospora crassa* mitochondria. *Eukaryotic Cell*, **6**, 2391-2405

Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, **322**, 1843-1845

Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*,

Massover, W.H., Lai, P.F., and Marsh, P. (2001) Negative staining permits 4.0 angstrom resolution with low-dose electron diffraction of catalase crystals. *Ultramicroscopy*, **90**, 7-12

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., Dekok, A., and Hol, W.G.J. (1992) Atomic-Structure of the Cubic Core of the Pyruvate-Dehydrogenase Multienzyme Complex. *Science*, **255**, 1544-1550

Minauro-Sanmiguel, F., Wilkens, S., and Garcia, J.J. (2005) Structure of dimeric mitochondrial ATP synthase: Novel F-0 bridging features and the structural basis of mitochondrial cristae biogenesis. *Proc. Natl. Acad. Sci. U. S. A*, **102**, 12356-12358

Mojica, F.J.M., ez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiol.*, **155**, 733-740

Mojica, F.J.M., ez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174-182

Monera, O.D., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J. Biol. Chem.*, **268**, 19218-19227

Nevo, R., Charuvi, D., Shimoni, E., Schwarz, R., Kaplan, A., Ohad, I., and Reich, Z. (2007) Thylakoid membrane perforations and connectivity enable intracellular traffic in cyanobacteria. *EMBO J.*, **26**, 1467-1473

Nicastro, D., Frangakis, A.S., Typke, D., and Baumeister, W. (2000) Cryo-electron tomography of *Neurospora* mitochondria. *J. Struc. Biol.*, **129**, 48-56

Oostergetel, G.T., Keegstra, W., and Brisson, A. (1998) Automation of specimen selection and data acquisition for protein electron crystallography. *Ultramicroscopy*, **74**, 47-59

Peters, K., Dudkina, N.V., Jänsch, L., Braun, H.P., and Boekema, E.J. (2008) A structural investigation of complex I and I+III_2 supercomplex from *Zea mays* at 11-13 angstrom resolution: Assignment of the carbonic anhydrase domain and evidence for structural heterogeneity within complex I. *Biochim. Biophys. Acta*, **1777**, 84-93

Pfeiffer, K., Gohil, V., Stuart, R.A., Hunte, C., Brandt, U., Greenberg, M.L., and Schägger, H. (2003) Cardiolipin stabilizes respiratory chain supercomplexes. *J. Biol. Chem.*, **278**, 52873-52880

Posten, C. and Schaub, G. (2009) Microalgae and terrestrial biomass as source for fuels-A process view. *J. Biotechnol.*, **142**, 64-69

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiol.*, **151**, 653-663

Prodöhl, A., Volkmer, T., Finger, C., and Schneider, D. (2005) Defining the structural basis for assembly of a transmembrane cytochrome. *J. Mol. Biol.*, **350**, 744-756

Prodöhl, A., Weber, M., Dreher, C., and Schneider, D. (2007) A mutational study of transmembrane helix-helix interactions. *Biochimie*, **89**, 1433-1437

Pulz, O. and Gross, W. (2004) Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.*, **65**, 635-648

Radermacher, M., Ruiz, T., Clason, T., Benjamin, S., Brandt, U., and Zickermann, V. (2006) The three-dimensional structure of complex I from *Yarrowia lipolytica*: A highly dynamic enzyme. *J. Struc. Biol.*, **154**, 269-279

Rich, P.R. (2003) The molecular machinery of Keilin's respiratory chain. *Biochem. Soc. Trans.*, **31**, 1095-1105

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979) Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *J. Gen. Microbiol.*, **111**, 1-61

Rossmann, M.G., Bernal, R., and Pletnev, S.V. (2001) Combining electron microscopic with X-ray crystallographic structures. *J. Struc. Biol.*, **136**, 190-200

Rupprecht, J., Hankamer, B., Mussgnug, J.H., Ananyev, G., Dismukes, C., and Kruse, O. (2006) Perspectives and advances of biological H_2 production in microorganisms. *Appl. Microbiol. Biotechnol.*, **72**, 442-449

Russell, D.G. (1995) Of Microbes and Macrophages - Entry, Survival and Persistence. *Cur. Opin. Immun.*, **7**, 479-484

Saraste, M. (1999) Oxidative phosphorylation at the *fin de siecle*. *Science*, **283**, 1488-1493

Sazanov, L.A. and Hinchliffe, P. (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science*, **311**, 1430-1436

Schäfer, E., Dencher, N.A., Vonck, J., and Parcej, D.N. (2007) Three-dimensional structure of the respiratory chain supercomplex $I_1III_2IV_1$ from bovine heart mitochondria. *Biochemistry*, **46**, 12579-12585

Schäfer, E., Seelert, H., Reifschneider, N.H., Krause, F., Dencher, N.A., and Vonck, J. (2006) Architecture of active mammalian respiratory chain supercomplexes. *J. Biol. Chem.*, **281**, 15370-15375

Schägger, H. and Pfeiffer, K. (2000) Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.*, **19**, 1777-1783

Schägger, H. and Pfeiffer, K. (2001) The ratio of oxidative phosphorylation complexes I-V in bovine heart mitochondria and the composition of respiratory chain supercomplexes. *J. Biol. Chem.*, **276**, 37861-37867

Schneider, D. and Engelman, D.M. (2003) GALLEX, a measurement of heterologous association of transmembrane helices in a biological membrane. *J. Biol. Chem.*, **278**, 3105-3111

Schneider, D. and Engelman, D.M. (2004a) Involvement of transmembrane domain interactions in signal transduction by alpha/beta integrins. *J. Biol. Chem.*, **279**, 9840-9846

Schneider, D. and Engelman, D.M. (2004b) Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J. Mol. Biol.*, **343**, 799-804

Schneider, D., Fuhrmann, E., Scholz, I., Hess, W., and Graumann, P. (2007) Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *Cell Biol.*, **8**, 39-

Schwartz, R.M. and Dayhoff, M.O. (1978) Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts. *Science*, **199**, 395-403

Semenova, E., Nagornykh, M., Pyatnitskiy, M., Artamonova, I.I., and Severinov, K. (2009) Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol. L.*, **296**, 110-116

Sheehan, J., Dunahay, T., and Benemann, J. (1998) A look back at the U.S. Department of Energy's Aquatic Species Program: biodiesel from algae. *National Renewable Energy Laboratory*,

Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.*, **6**, 181-186

Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, **287**, 252-260

Standar, K., Mehner, D., Osadnik, H., Berthelmann, F., Hause, G., Lünsdorf, H., and Brüser, T. (2008) PspA can form large scaffolds in *Escherichia coli*. *FEBS Letters*, **582**, 3585-3589

Stock, D., Leslie, A.G.W., and Walker, J.E. (1999) Molecular architecture of the rotary motor in ATP synthase. *Science*, **286**, 1700-1705

Strauss, M., Höfhaus, G., Schröder R.R., and Kühlbrandt W. (2008) Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. *EMBO J.*, **27**, 1154-1160

Sunderhaus, S., Dudkina, N.V., Jänsch, L., Klodmann, J., Heinemeyer, J., Perales, M., Zabaleta, E., Boekema, E.J., and Braun, H.P. (2006) Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *J. Biol. Chem.*, **281**, 6482-6488

Svergun, D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J.Appl.Cryst.*, **25**, 495-503

Tahallah, N., Pinkse, M., Maier, C.S., and Heck, A.J.R. (2001) The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun. Mass Spectrom.*, **15**, 596-601

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U. S. A*, **99**, 7536-7541

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl.Acids Res.*, **25**, 4876-4882

Tice, M.M. and Lowe, D.R. (2006) Hydrogen-based carbon fixation in the earliest known photosynthetic organisms. *Geology*, **34**, 37-40

Tzinas, G., Argyroudi-Akoyunoglou, J.H., and Akoyunoglou, G. (1987) The Effect of the Dark Interval in Intermittent Light on Thylakoid Development - Photosynthetic Unit Formation and Light Harvesting Protein Accumulation. *Photosyn. Res.*, **14**, 241-258

van de Meene, A.M.L., Hohmann-Marriott, M.F., Vermaas, W.F.J., and Roberson, R.W. (2006) The three-dimensional structure of the cyanobacterium *Synechocystis* sp PCC 6803. *Arch. Microbiol.*, **184**, 259-270

van den Heuvel, R.H.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J.R. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.*, **78**, 7473-7483

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sc.*, **34**, 401-407

van der Ploeg, J.R. (2009) Analysis of CRISPR in *Streptococcus* mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiol.*, **155**, 1966-1976

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Quart. Rev. Biophys.*, **33**, 307-369

Volkov, V.V. and Svergun, D.I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J.Appl.Cryst.*, **36**, 860-864

Vothknecht, U.C. and Soll, J. (2005) Chloroplast membrane transport: interplay of prokaryotic and eukaryotic traits. *Gene*, **354**, 99-109

Vothknecht, U.C. and Westhoff, P. (2001) Biogenesis and origin of thylakoid membranes. *Biochim. Biophys. Acta*, **1541**, 91-101

Waghmare, S.P., Pousinis, P., Hornby, D.P., and Dickman, M.J. (2009) Studying the mechanism of RNA separations using RNA chromatography and its application in the analysis of ribosomal RNA and RNA:RNA interactions. *J. Chromatogr.*, **1216**, 1377-1382

Weiner, L. and Model, P. (1994) Role of An *Escherichia coli* Stress-Response Operon in Stationary-Phase Survival. *Proc. Natl. Acad. Sci. U. S. A*, **91**, 2191-2195

Westphal, S., Heins, L., Soll, J., and Vothknecht, U.C. (2001) Vipp1 deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? *Proc. Natl. Acad. Sci. U. S. A*, **98**, 4243-4248

Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucl.Acids Res.*, **32**, 668-673

Whitmore, L. and Wallace, B.A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*, **89**, 392-400

Wiedenheft, B., Zhou, K.H., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009) Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense. *Structure*, **17**, 904-912

Wittig, I., Braun, H.P., and Schagger, H. (2006a) Blue native PAGE. *Nat. Prot.*, **1**, 418-428

Wittig, I., Carrozzo, R., Santorelli, F.M., and Schägger, H. (2006b) Supercomplexes and subcomplexes of mitochondrial oxidative phosphorylation. *Biochim. Biophys. Acta*, **1757**, 1066-1072

Wittig, I. and Schägger, H. (2008) Structural organization of mitochondrial ATP synthase. *Biochim. Biophys. Acta*, **1777**, 592-598

Wriggers, W., Milligan, R.A., and McCammon, J.A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struc. Biol.*, **125**, 185-195

Yamaguchi, M., Danev, R., Nishlyama, K., Sugawara, K., and Nagayama, K. (2008) Zernike phase contrast electron microscopy of ice-embedded influenza A virus. *J. Struc. Biol.*, **162**, 271-276

Yeremenko, N., Kouril, R., Ihalainen, J.A., D'Haene, S., van Oosterwijk, N., Andrizhiyevskaya, E.G., Keegstra, W., Dekker, H.L., Hagemann, M., Boekema, E.J., Matthijs, H.C.P., and Dekker, J.P. (2004) Supramolecular Organization and Dual Function of the IsiA Chlorophyll-Binding Protein in Cyanobacteria. *Biochemistry*, **43**, 10308-10313

Yu, B., Yang, Z.Y., Li, J.J., Minakhina, S., Yang, M.C., Padgett, R.W., Steward, R., and Chen, X.M. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932-935

Zhang, M., Mileykovskaya, E., and Dowhan, W. (2002) Gluing the respiratory chain together - Cardiolipin is required for supercomplex formation in the inner mitochondrial membrane. *J. Biol. Chem.*, **277**, 43553-43556

Zhang, X., Jin, L., Fang, Q., Hui, W.H., and Zhou, Z.H. (2010) 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell*, **141**, 472-482

Zhou, M., Sandercock, A.M., Fraser, C.S., Ridlova, G., Stephens, E., Schenauer, M.R., Yokoi-Fong, T., Barsky, D., Leary, J.A., Hershey, J.W., Doudna, J.A., and Robinson, C.V. (2008) Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc. Natl. Acad. Sci. U. S. A*, **105**, 18139-18144

Co-author affiliations

Egbert J. Boekema, Roman Kouřil, Department of Biophysical Chemistry, GBB, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Hans-Peter Braun, Institute for Plant Genetics, Faculty of Natural Sciences, Universität Hannover, Herren-häuser Straße 2, 30419 Hannover, Germany

Marieke R. Beijer, Stan J.J. Brouns, Matthijs M. Jore, Magnus Lundgren[a], John van der Oost, Edze R. Westra, Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

Arjan Barendregt, Esther van Duijn, Albert J. R. Heck, Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research, Utrecht Institute for Pharmaceutical Sciences, and The Netherlands Proteomics Center, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Mark J. Dickman, Ambrosius P.L. Snijders, Sakharam P. Waghmare, ChELSI Institute, Department of Chemical and Process Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Jennifer A. Doudna, Blake Wiedenheft, Kaihong Zhou, Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

Eva Fuhrmann[b], Eva Rupprecht, Dirk Schneider[b], Institut für Biochemie und Molekularbiologie, ZBMZ, Stefan-Meier-Straße 17, D-79104 Freiburg, Germany

Uwe Kahmann, Fakultät für Biologie, Universität Bielefeld, D-33501 Bielefeld, Germany

Present address

[a.] Department of Cell and Molecular Biology, Uppsala University, SE-751 24 Uppsala, Sweden

[b.] Institut für Pharmazie und Biochemie, Johannes Gutenberg-Universität Mainz, Johann-Joachim-Becher-Weg 30, 55128 Mainz, Germany

In een notendop

In een notendop

Het onderzoek dat beschreven is in dit proefschrift is voor een groot gedeelte tot stand gekomen door gebruik te maken van één instrument (de transmissie elektronenmicroscoop), één praktische handeling (negatief kleuren) en één verwerkingsmethode (losse-deeltjes-middeling). In het volgende gedeelte worden deze drie facetten nader toegelicht.

De transmissie elektronenmicroscoop

De transmissie elektronenmicroscoop (TEM) is in essentie niet veel anders dan een lichtmicroscoop (LM). Beide microscopen bestaan uit een kolom met aan één uiteinde een lichtbron en aan de andere kant een detector. Het licht van deze bron vormt een bundel die beïnvloed wordt door een aantal diafragma's, meerdere vergrotingslenzen en het te bekijken object (monster). Het beeld dat van het monster wordt gevormd komt uiteindelijk terecht bij een detector. Voor de LM kan dit een filmplaat, CCD-camera of het menselijke oog zijn. Het laatstgenoemde is praktisch gezien niet mogelijk bij de TEM en dat is maar goed ook, omdat het oog daardoor beschadigd kan worden. Dit heeft te maken met het soort lichtbundel waarvan gebruik wordt gemaakt bij de TEM, namelijk elektronen. Bij de LM is dit daarentegen gewoon zichtbaar licht.

Het verschil van licht heeft gevolgen voor de opbouw van de beide soorten microscopen. Het meest zichtbare verschil is de grootte van de microscoop: een LM past prima op een tafel en is doorgaans niet groter dan ~50 cm. De TEM, echter, torent met gemak boven iedereen uit. Verder heerst er in de kolom van de TEM een hoog vacuüm. Hierdoor gedragen elektronen zich net zo als lichtstralen in lucht. De kans is namelijk nagenoeg nul dat een elektron, die zich door een vacuüm voortbeweegt, in botsing komt met een luchtmolecuul. Hierdoor verandert de richting van een elektron alleen door interacties met het monster en het lenzensysteem.

Deze lenzen vormen een ander verschil tussen de TEM en de LM. In de LM zijn de lenzen gemaakt van glas of een soortgelijk doorzichtig materiaal, in de TEM worden elektromagneten gebruikt als lenzen. Door de stroom op de elektromagneten te variëren verandert het magnetisch veld en wordt zodoende de vergroting van deze lenzen ook sterker of zwakker.

Het grote voordeel van de TEM ten opzichte van de LM is dat de golflengte van elektronen veel kleiner is dan die van zichtbaar licht. De golflengte van licht varieert van

0,0004 tot 0,0008 millimeter. Dit betekent dat objecten die kleiner zijn dan die golflengte, zoals eiwitcomplexen, niet waarneembaar zijn. De golflengte van elektronen is zo klein dat je individuele atomen kunt waarnemen. Dankzij de voortdurende technologische ontwikkeling van elektronenmicroscopen is dit ook bijna mogelijk geworden voor biologische objecten als eiwitten. De methode gebruikt voor dit onderzoek maakt details tot ongeveer 0,000015 millimeter zichtbaar.

Negatief kleuren

Het negatief kleuren, de manier van monsterpreparatie die veelvuldig is toegepast tijdens dit promotieonderzoek, wordt hier verder uitgelegd. Om negatief te kleuren wordt een "gridje", een metalen plaatje met een diameter van ~2 mm en een heleboel kleine openingen, bedekt met een heel dun vliesje van koolstof. Dit koolstofvliesje is dun genoeg om er met behulp van de TEM door heen te kijken maar wel zo sterk dat het de te onderzoeken eiwitten kan dragen.

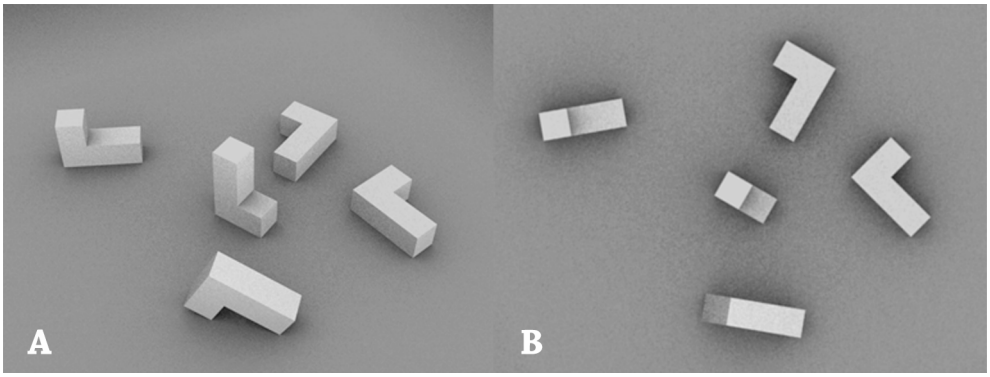
Het lastige van eiwitten is dat ze voor een groot gedeelte ook uit koolstof bestaan en hierdoor, zonder verdere behandeling, amper te onderscheiden zijn van het koolstofvliesje. Om toch onderscheid te kunnen maken tussen de eiwitten en het koolstofvliesje wordt het geheel behandeld met vloeibaar kleurmiddel. De truck van de hele preparatie is om precies zoveel kleurmiddel weg te halen dat alleen rondom de eiwitten nog kleurmiddel achterblijft.

Zoals eerder al is besproken wordt bij de TEM gebruik gemaakt van elektronen. De richting van de elektronen wordt beïnvloed als ze door het met kleurmiddel geprepareerde monster gaan. De mate van beïnvloeding hangt vooral af van de atomen die de elektronen onderweg tegenkomen. Als deze atomen relatief licht zijn, zoals bij het koolstofvlies en de eiwitten, dan worden de elektronen amper beïnvloed. Dit betekent dat ze rechtdoor gaan en in de uiteindelijke opname een lichte plek veroorzaken. Zijn de atomen zwaar, zoals die van het kleurmiddel, dan worden ze wel beïnvloed en verandert de baan van de elektronen dusdanig dat ze niet bij de detector terechtkomen. Dit resulteert in een donkere plek in de uiteindelijke opname. Dit is de truck van negatief kleuren: de omtrek van het eiwit wordt zichtbaar gemaakt en niet het eiwit zelf.

Losse-deeltjes-middeling

De beeldverwerking van losse deeltjes is het eenvoudigst uit te leggen aan de hand van het L-vormig deeltje. Dit 3-dimensionaal deeltje kan, zoals weergegeven in figuur 1a, op verschillende manieren terechtkomen op een oppervlak. Van een preparaat met L-

vormige deeltjes worden vervolgens foto's genomen. Normaal gesproken staat de lichtbundel loodrecht op het preparaat en resulteert dit in een opname zoals weergegeven is in figuur 1b.

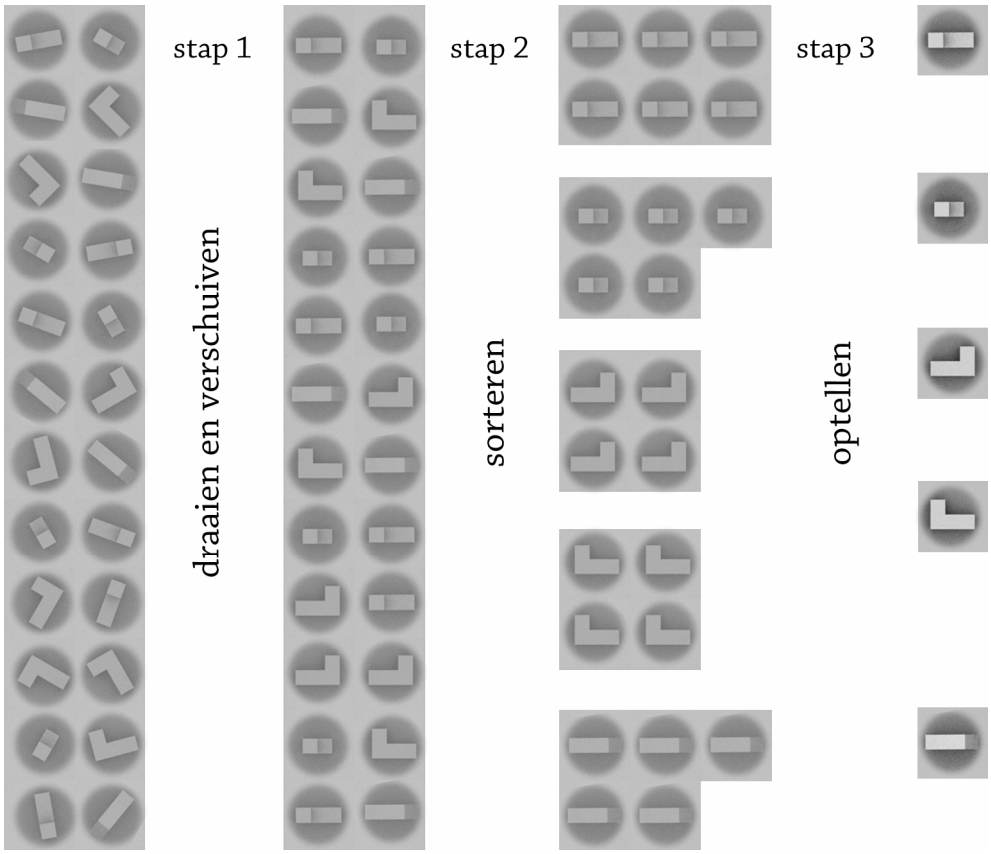


Figuur 1. Het L-vormig deeltje. In (A) zijn de vijf verschillende manieren weergegeven (in perspectief) hoe een L-vormig deeltje op een oppervlak terecht kan komen. (B) Dezelfde vijf oriëntaties als in A alleen nu loodrecht gezien op het oppervlak.

In tegenstelling tot wat te zien is in bovenstaand figuur zijn eiwitten meestal niet erg duidelijk te onderscheiden van de achtergrond. Daarom worden per monster meestal meerdere (honderden) opnames genomen. Tegenwoordig kan dit semi-automatisch worden gedaan. Met behulp van Grace, een programma ontwikkelt door Gert Oostergetel, is het mogelijk om aan te geven waar de microscoop foto's van het preparaat moet maken. In de praktijk komt het erop neer dat overdag een mooi preparaat wordt gemaakt waarna de microscoop m.b.v. Grace wordt geïnstrueerd om gedurende de nacht opnames van het preparaat te maken.

De volgende dag kan worden begonnen met de beeldverwerking waarvoor Wilko Keegstra (net als Gert ook werkzaam binnen de EM-vakgroep van Groningen) het softwarepakket GRIP (GRoningen Image Processing) heeft ontwikkeld. Uit elke opname worden kleine vierkante stukjes gesneden met in het midden één deeltje (figuur 2). Dit kan semi-automatisch worden gedaan waardoor met gemak een dataset ontstaat met duizenden losse deeltjes. Door deze losse deeltjes zo te draaien en te verschuiven (stap 1) ontstaat een aangepaste dataset waarbij alle losse deeltjes op dezelfde manier liggen. De verschillende aanzichten moeten van elkaar worden gescheiden. Dit kan d.m.v. een aantal langdurige berekeningen die hier verder niet worden toegelicht (stap 2). De laatste stap (stap 3) is het optellen van al deze losse deeltjes wat resulteert in één gemiddelde

projectie. In deze uiteindelijke projectie worden de vorm en details van het deeltje duidelijk zichtbaar, de achtergrond verliest juist vorm en wordt egaal.



Figuur 2. Het beeldverwerkingsproces weergegeven vanaf het begin (een onbewerkte dataset met losse deeltjes) tot en met de uiteindelijke gemiddelde projectie. Het proces omvat drie stappen. Tijdens de eerst stap worden de losse deeltjes (m.b.v. software) zo gedraaid dat deeltjes die op dezelfde manier op een oppervlak liggen ook een gelijke onderlinge oriëntatie hebben. Tijdens de tweede stap worden de verschillende deeltjes (qua ligging op het oppervlak) gesorteerd. Tot slot worden alle deeltjes binnen één groep opgeteld zodat de verhouding tussen het signaal en de ruis verbeterd wordt en een gemiddelde van elke groep wordt verkregen.

Het L-vormig deeltje komt ook in de natuur voor. Heel veel organismen, van klein tot groot, hebben een eiwitcomplex dat complex I wordt genoemd. In figuur 3 van hoofdstuk 2 (blz. 41, dit proefschrift) staan verschillende aanzichten van complex I. Dit complex is samen met drie andere eiwitcomplexen betrokken bij de productie van energie uit voedsel. Drie van deze vier eiwitcomplexen (complex I, III en IV) kunnen samen een keten vormen in het membraan van mitochondrieën. Het in handen krijgen van deze keten, om er bijvoorbeeld met de TEM naar te kijken, is niet gemakkelijk.

Om een idee te krijgen van hoe de keten van eiwitcomplexen er uitziet hebben we fragmenten van deze keten bekeken en geanalyseerd. Het resultaat hiervan is te zien in figuur 4 en 5 van hoofdstuk 2 (blz. 42-43, dit proefschrift). Hierin zijn verschillende fragmenten te zien die bestaan uit bijvoorbeeld complex I en complex III, of complex I, complex III en complex IV. Omdat er van deze ketenfragmenten verschillende aanzichten beschikbaar zijn is het mogelijk de onderlinge verbanden te bepalen, zie figuur 6 (blz. 46, dit proefschrift). Met behulp van de resultaten van anderen is het mogelijk de keten te herbouwen. Dit schematische model, te zien in figuur 7 (blz. 49, dit proefschrift), voldoet aan alle beschikbare data, zowel aan die van ons als van anderen. De toekomst zal moeten uitwijzen of de voorgestelde structuur van de keten er ook daadwerkelijk zo uitziet.

Dankwoord

Dankwoord

Maandagavond, de laptop is aan, de koffie alweer op en nu moet het toch echt gaan gebeuren: het dankwoord. Niet dat er geen mensen te bedanken zijn, dat is zeker niet het geval. Maar nadat alle hoofdstukken eindelijk af zijn, qua inhoud, lay-out, paginanummering, alle beestjes en plantjes *cursief*, etcetera, is het proefschrift nog niet klaar voor de drukker. Van alles heb ik dit gedeelte het langst uitgesteld, probeer maar eens origineel te zijn. De tweede kop koffie staat inmiddels naast me. Er is eigenlijk nog niet zoveel veranderd sinds daarnet, het echte dankwoord moet nog steeds geschreven worden. Het liefst natuurlijk niet in de vorm van een lange opsomming van collega's, oud-collega's, mensen van de andere vakgroep en daar dan ook de mensen van die alweer zijn vertrokken. Dat zou op zich allemaal nog wel te doen zijn, maar dan heb je nog de tijdelijke onderzoekers die twee of drie maanden meedraaien in het lab en de studenten van andere promovendi. Zelf heb ik geen studenten begeleid tijdens de afgelopen vier jaar, alleen een aantal practica gegeven voor grotere groepen studenten, maar dat terzijde. Dan zijn er nog de mensen buiten Groningen. Ik heb samengewerkt met mensen uit verschillende steden in Nederland, Duitsland, Groot-Brittanië en de Verenigde Staten. Conclusie: dat zijn er heel wat. Daarom ga ik proberen het anders te doen en zie ik wel hoe het uitpakt.

Om te beginnen bedank ik alle mensen van de Elektronen Microscopie Groep. Jullie hebben mij de kneepjes van het vak geleerd, waren altijd in voor een praatje over het werk of andere dingen en nooit te beroerd om mij te helpen. Bedankt! Ook ben ik de mensen van de Eiwit Kristallografie Groep dankbaar. Als student heb ik al een leuke tijd bij jullie gehad en dat is daarna als buurman altijd zo gebleven. Beter een goede buur… . Bedankt!

A special thank to all the people that collaborated with me. You prepared, and, more important, shared with me the most beautiful samples. And as it turned out, beautiful they are. Thanks again!

Ik besef me net dat ik met deze aanpak een heel eind kom, alleen familie en vrienden nog bedanken en klaar is het dankwoord. Maar dat gaat niet gebeuren, er zijn een aantal mensen die ik specifiek wil noemen. Marcel en Niels, allereerst bedankt dat jullie mijn paranimfen willen zijn en dat ik in de nabije toekomst ook jullie paranimf mag zijn. We kennen elkaar al een kleine tien jaar en hebben sinds het derde jaar van de HLO tot nu ongeveer hetzelfde traject gevolgd: een jaar stage gelopen bij de RuG, vervolgens Scheikunde gestudeerd aan diezelfde RuG en daarna zijn we samen op één kamer

terechtgekomen om te promoveren, nog steeds bij de RuG. Tijdens al die jaren heb ik ontzettend veel lol met jullie gehad, maar ook serieuze gesprekken gevoerd, veel te veel bier gedronken, spelletjes gespeeld, steun en inspiratie gehad en nog wel meer. Waar we in de toekomst ook terechtkomen en wat we ook gaan doen, deze tijd hebben we mooi te pakken en die zal ik nooit vergeten. Ontzettend bedankt!

Roman, the first two years of my PhD period, you were my daily supervisor. During this time you taught me what electron microscopy and science in practice was. The last two years we didn't share the same project anymore but you always stayed involved in what I was doing. Your involvement in my PhD period, patience in explaining me all kind of things and stimulation to just do things instead of complaining that there are many things that I still have to do, this all made me the scientist that I am today. A part from all these things you are also a very nice person to drink Czech beers and discuss al kind of topics. Altogether, thank you very much!

Lieve papa en mama, zonder jullie steun was ik nooit zover gekomen. Ik heb het vermoeden dat hoe langer ik aan het studeren was, des te ingewikkelder het voor jullie begon te worden. Toch zijn jullie altijd geïnteresseerd en betrokken geweest bij wat ik de afgelopen vier jaar heb gedaan. Dit waardeer ik enorm. Bedankt daarvoor! En Heine, mijn grote broer. Ik heb nog altijd veel respect voor je keuze van toen om naar Den Haag te gaan. Josephien en ik vinden het al moeilijk om überhaupt over verhuizen na te denken, maar jij bent toen gewoon gegaan. En telkens als we langskwamen kreeg ik niet de indruk dat je daar spijt van hebt gekregen. Aan zo'n sprong in het diepe kan ik nog een voorbeeld nemen. Ik ben er trots op dat je mijn broer bent.

Tot slot wil ik jou, Josephien, bedanken. Je hebt bijna mijn hele periode als promovendus meegemaakt, eerst elk vanuit onze eigen kamers en daarna, wat gelukkig nog steeds zo is, vanuit ons huisje. Ik heb veel bewondering voor hoe jij mij mijn gang liet gaan als ik toch nog even naar het lab wilde. Die keren dat je gezellig met me mee ging, ik achter de microscoop en jij spelend met stikstof of een krantje lezend met een kopje thee erbij, vond ik erg fijn. Ik hoop dat jij nog heel lang in mijn leven blijft.

Jelle

List of publications

List of publications

Thangaraj, B., Jolley, C., **Bultema J.B.**, Whitelegge, J., Lin, S., Kouřil, R., Greyslak, J., Boekema, E.J. & Fromme P. Efficiency at the expense of flexibility: The structure and function of PSI-LHCI from *Galdieria sulphuraria*. In preparation.

Jore, M.M.*, Lundgren, M.*, van Duijn, E.*, **Bultema, J.B.***, Westra, E.R., Waghmare, S.P., Wiedenheft, B., Beijer, M.R., Barendregt A., Zhou, K., Snijders, A.P.L., Dickman, M.J., Doudna, J., Boekema, E.J., Heck, A.J.R., van der Oost J. & Brouns, S.J.J. Structural basis for CRISPR RNA-guided recognition of DNA by Cascade. In preparation.

Bultema, J.B., Fuhrmann, E., Boekema E.J. & Schneider D. (2010) Vipp1 and PspA: related but no twins. *Communicative and Integrative Biology*. Accepted.

Dudkina, N.V., Kouřil, R., **Bultema, J.B.** & Boekema, E.J. (2010) Imaging of organelles by electron microscopy reveals protein-protein interactions in mitochondria and chloroplasts. *FEBS Letters*. Accepted.

Bultema, J.B., Braun, H.P., Boekema, E.J. & Kouřil, R. (2009) Megacomplex organization of the oxidative phosphorylation system by structural analysis of respiratory supercomplexes from potato. *Biochim.Biophys.Acta*, 1787, p. 60-67.

Fuhrmann, E., **Bultema, J.B.**, Kahmann, U., Rupprecht, E., Boekema, E.J. & Schneider, D. (2009) The Vesicle-inducing Protein 1 from Synechocystis sp. PCC 6803 Organizes into Diverse Higher-Ordered Ring Structures. *Molecular Biology of the Cell*, 20, p. 4620-4628.

Levisson, M., Sun, L., Hendriks, S., Swinkels, P., Akveld, T., **Bultema, J.B.**, Barendregt, A., van den Heuvel, R.H.H., Dijkstra, B.W., van der Oost, J. & Kengen, S.W.M. (2009) Crystal Structure and Biochemical Properties of a Novel Thermostable Esterase Containing an Immunoglobulin-Like Domain. *J.Mol.Biol.*, 385, p. 949-962.

* These authors contributed equally to this work